Robust Signaling for Bursty Interference

Grace Villacrés [1,2,*][ID], **Tobias Koch** [1,2][ID], **Aydin Sezgin** [3][ID] **and Gonzalo Vazquez-Vilar** [1,2][ID]

1   Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911 Leganés, Spain;
     koch@tsc.uc3m.es (T.K.); gonzalo.vazquez@uc3m.es (G.V.-V.)
2   Gregorio Marañón Health Research Institute, 28007 Madrid, Spain
3   Electrical Engineering and Information Technology Department, Ruhr-Universität Bochum,
     44780 Bochum, Germany; aydin.sezgin@rub.de
*   Correspondence: gvillacres@tsc.uc3m.es; Tel: +34-916-248-839.

Abstract: This paper studies a bursty interference channel, where the presence/absence of interference is modeled by a block-i.i.d. Bernoulli process that stays constant for a duration of T symbols (referred to as coherence block) and then changes independently to a new state. We consider both a quasi-static setup, where the interference state remains constant during the whole transmission of the codeword, and an ergodic setup, where a codeword spans several coherence blocks. For the quasi-static setup, we study the largest rate of a coding strategy that provides reliable communication at a basic rate and allows an increased (opportunistic) rate when there is no interference. For the ergodic setup, we study the largest achievable rate. We study how non-causal knowledge of the interference state, referred to as *channel-state information* (*CSI*), affects the achievable rates. We derive converse and achievability bounds for (i) local CSI at the receiver side only; (ii) local CSI at the transmitter and receiver side; and (iii) global CSI at all nodes. Our bounds allow us to identify when interference burstiness is beneficial and in which scenarios global CSI outperforms local CSI. The joint treatment of the quasi-static and ergodic setup further allows for a thorough comparison of these two setups.

Keywords: bursty interference channel; channel-state information; linear deterministic model; ergodic case; quasi-static case; sum capacity; opportunistic rates

1. Introduction

Interference is a key limiting factor for the efficient use of the spectrum in modern wireless networks. It is, therefore, not surprising that the *interference channel* (*IC*) has been studied extensively in the past; see, e.g., [1] (Chapter 6) and references therein. Most of the information-theoretic work developed for the IC assumes that interference is always present. However, certain physical phenomena, such as shadowing, can make the presence of interference intermittent or bursty. Interference can also be bursty due to the bursty nature of data traffic, distributed medium access control mechanisms, and decentralized networking protocols. For this reason, there has been an increasing interest in understanding and exploring the effects of burstiness of interference.

Seminal works in this area were performed by Khude et al. in [2] for the Gaussian channel and in [3] by using a model which corresponds to an approximation to the two-user Gaussian IC. They tried to harness the burstiness of the interference by taking advantage of the time instants when the interference is not present to send opportunistic data. Specifically, [2,3] considered a channel model where the interference state stays constant during the transmission of the entire codeword, which corresponds to a quasi-static channel. Motivated by the idea of degraded message sets by Körner and Marton [4], Khude et al. studied the largest rate of a coding strategy that provides reliable communication at a basic rate R and allows an increased (opportunistic) rate $R + \Delta R$ when there is no

interference. The idea of opportunism was also used by Diggavi and Tse [5] for the quasi-static flat fading channel and, recently, by Yi and Sun [6] for the *K*-user IC with states.

Wang et al. [7] modeled the presence of interference using an *independent and identically distributed* (*i.i.d.*) Bernoulli process that indicates whether interference is present or not, which corresponds to an ergodic channel. They further assume that the interference links are fully correlated. Wang et al. mainly studied the effect of causal feedback under this model, but also presented converse bounds for the non-feedback case. Mishra et al. considered the generalization of this model to multicarrier systems, modeled as parallel two-user bursty ICs, for the feedback [8] and non-feedback case [9].

The bursty IC is related to the binary fading IC, for which the four channel coefficients are in the binary field $\{0, 1\}$ according to some Bernoulli distribution. Note, however, that neither of the two models is a special case of the other. While a zero channel coefficient of the cross link corresponds to intermittence of interference, the bursty IC allows for non-binary signals. Conversely, in contrast to the binary fading IC, the direct links in the bursty IC cannot be zero, since only the interference can be intermittent. Vahid et al. [10–14] studied the capacity region of the binary fading IC. Specifically, [11,14] study the capacity region of the binary fading IC when the transmitters do not have access to the channel coefficients, and [12] study the capacity region when the transmitters have access to the past channel coefficients. Vahid and Calderbank additionally study the effect on the capacity region when certain correlation is available to all nodes as side information [13].

The focus of the works by Khude et al. [3] and Wang et al. [7] was on the *linear deterministic model* (*LDM*), which was first introduced by Avestimehr [15], but falls within the class of more general deterministic channels whose capacity was obtained by El Gamal and Costa in [16]. The LDM maps the Gaussian IC to a channel whose outputs are deterministic functions of their inputs. Bresler and Tse demonstrated in [17] that the generalized degrees of freedom (first-order capacity approximation) of the two-user Gaussian IC coincides with the normalized capacity of the corresponding deterministic channel. The LDM thus offers insights on the Gaussian IC.

1.1. Contributions

In this work, we consider the LDM of a bursty IC. We study how interference burstiness and the knowledge of the interference states (throughout referred to as *channel-state information* (*CSI*)) affects the capacity of this channel. We point out that this CSI is different from the one sometimes considered in the analysis of ICs (see, e.g., [18]), where CSI refers to knowledge of the channel coefficients. (In this regard, we assume that all transmitters and receivers have access to the channel coefficients). For the sake of compactness, we focus on non-causal CSI and leave other CSI scenarios, such as causal or delayed CSI, for future work.

We consider the following cases: (i) only the receivers know the corresponding interference state (local CSIR); (ii) transmitters and receivers know their corresponding interference states (local CSIRT); and (iii) both transmitters and receivers know all interference states (global CSIRT). For each CSI level we consider both (i) the quasi-static channel and (ii) the ergodic channel. Specifically, in the quasi-static channel the interference is present or absent during the whole message transmission and we harness the realizations when the channel experiences better conditions (no presence of interference) to send extra messages. In the ergodic channel the presence/absence of interference is modeled as a Bernoulli random variable which determines the interference state. The interference state stays constant for a certain coherence time *T* and then changes independently to a new state. This model includes the i.i.d. model by Wang et al. as a special case, but also allows for scenarios where the interference state changes more slowly. Note, however, that when the receivers know the interference state (as we shall assume in this work), then the capacity of this model becomes independent of *T* and coincides with that of the i.i.d. model. The proposed analysis is performed for the two extreme cases where the states of each of the interfering links are independent, and where states of the interfering links are fully correlated. Hence we unify the scenarios already treated in the literature [2,3,7]. Nevertheless, some of our presented results can be extended to consider an arbitrary correlation between the interfering

states. The works by Vahid and Calderbank [13] and Yeh and Wang [19] characterize the capacity region of the two-user binary IC and the MIMO X-channel, respectively. While [13,19] consider a general spatial correlation between communication and interfering links, they do not consider the correlation between interfering links.

Our analysis shows that, for both the quasi-static and ergodic channels, for all interference regions except the very strong interference region, global CSIRT outperforms local CSIR/CSIRT. This result does not depend on the correlation between the states of the interfering links. For local CSIR/CSIRT and the quasi-static scenario, the burstiness of the channel is of benefit only in the very weak and weak interference regions. For the ergodic case and local CSIR, interference burstiness is only of clear benefit if the interference is either weak or very weak, or if it is present at most half of the time. This is in contrast to local CSIRT, where interference burstiness is beneficial in all interference regions.

Specific contributions of our paper include:

- A joint treatment of the quasi-static and the ergodic model: Previous literature on the bursty IC considers either the quasi-static model or the ergodic model. Furthermore, due to space constraints, the proofs of some of the existing results were either omitted or contain little details. In contrast, our paper discusses both models, allowing for a thorough comparison between the two.
- Novel achievability and converse bounds: For the ergodic model, the achievability bounds for local CSIRT, and the achievability and converse bounds for global CSIRT, are novel. In particular, novel achievability strategies are proposed that exploit certain synchronization between the users. To keep the paper self-contained, we further present the proof of the achievability bound for local CSIR that has appeared in the literature without proof.
- Novel converse proofs for the quasi-static model: In contrast to existing converse bounds, which are based on Fano's inequality, our proofs of the converse bounds for the rates of the worst-case and opportunistic messages are based on an information density approach (more precise, they are based on the Verdú-Han lemma). This approach does not only allow for rigorous yet clear proofs, but it would also enable a more refined analysis of the probabilities that worst-case and opportunistic messages can be decoded correctly.
- A thorough comparison of the sum capacity of various scenarios: *Inter alia*, the obtained results are used to study the advantage of featuring different levels of CSI, the impact of the burstiness of the interference, and the effect of the correlation between the channel states of both users.

The rest of this paper is organized as follows. Section 2 introduces the system model, where we define the bursty IC quasi-static setup, the ergodic setup, and briefly summarize previous results on the non-bursty IC. In Sections 3–5 we present our results for local CSIR, local CSIRT and global CSIRT, respectively. Section 6 studies the impact of featuring different CSI levels. Section 7 analyzes in which scenarios exploiting burstiness of interference is beneficial. Section 8 concludes the paper with a summary of the results. Most proofs of the presented results are deferred to the appendix.

1.2. Notation

To differentiate between scalars, vectors, and matrices we use different fonts: scalar random variables and their realizations are denoted by upper and lower case letters, respectively, e.g., B, b; vectors are denoted using bold face, e.g., $\boldsymbol{X}$, $\boldsymbol{x}$; random matrices are denoted via a special font, e.g., $\mathbb{X}$; and for deterministic matrices we shall use yet another font, e.g., S. For sets we use the calligraphic font, e.g., $\mathcal{S}$. We denote sequences such as $A_{i,1}, \ldots, A_{i,M}$ by A_i^M. We define $\max\{0, x\}$ as $(x)^+$.

We use $\mathbb{F}_2$ to denote the binary Galois field and $\oplus$ to denote the modulo 2 addition. Let the down-shift matrix $\mathsf{S}_u \in \mathbb{F}_2^{q \times q}$, a matrix of dimension $q \times q$, be defined as

$$\mathsf{S}_u = \begin{bmatrix} 0_{u \times (q-u)}^T & 0 \\ \mathsf{I}_u & 0_{u \times (q-u)} \end{bmatrix}_{q \times q}$$

with $0_{q-1} \in \mathbb{F}_2^{q-1}$ the all-zero vector and $I_u \in \mathbb{F}_2^{u \times u}$ the identity matrix.

Similarly, we define the matrix $L_d \in \mathbb{F}_2^{q \times q}$ of dimension $q \times q$ that selects the d lowest components of a vector of dimension q:

$$L_d = \begin{bmatrix} 0 & 0_{d \times (q-d)}^T \\ 0_{d \times (q-d)} & I_d \end{bmatrix}_{q \times q}.$$

We shall denote by $H_b(p)$ the entropy of a binary random variable X with probability mass function $(p, 1-p)$, i.e.,

$$H_b(p) \triangleq -p \log p - (1-p) \log(1-p). \tag{1}$$

Similarly, we denote by $H_{\text{sum}}(p, q)$ the entropy $H(X \oplus \tilde{X})$ where X and $\tilde{X}$ are two independent binary random variables with probability mass functions $(p, 1-p)$ and $(q, 1-q)$, respectively:

$$lClH_{\text{sum}}(p, q) \triangleq H_b(p(1-q) + (1-p)q) \tag{2}$$

For this function it holds that $H_{\text{sum}}(p, q) = H_{\text{sum}}(1-p, q) = H_{\text{sum}}(p, 1-q) = H_{\text{sum}}(1-p, 1-q)$. Finally, $\mathbb{1}(\cdot)$ denotes the indicator function, i.e., $\mathbb{1}(\text{statement})$ is 1 if the statement is true and 0 if it is false.

2. System Model

Our analysis is based on the LDM, introduced by Avestimehr et al. [15] for some relay network. This model is, on the one hand, simple to analyze and, on the other hand, captures the essential structure of the Gaussian channel in the high signal-to-noise ratio regime.

We consider a bursty IC where (i) the interference state remains constant during the whole transmission of the codeword of length N (quasi-static setup) or (ii) the interference state remains constant for a duration of T consecutive symbols and then changes independently to a new state (ergodic setup). For one coherence block, the two-user bursty IC is depicted in Figure 1, where n_d and n_c are the channel gains of the direct and cross links, respectively. We assume that n_d and n_c are known to both the transmitter and receiver and remain constant during the whole transmission of the codeword. For simplicity, we shall assume that n_d and n_c are equal for both users. Nevertheless, most of our results generalize to the asymmetric case. More precisely, all converse and achievability bounds generalize to the asymmetric case, while the direct generalization of the proposed achievability schemes may be loose in some asymmetric regions.
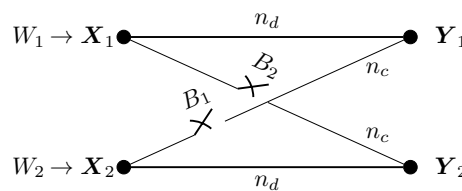

Figure 1. Channel model of the bursty interference channel.

For the k-th block, the input-output relation of the channel is given by

$$Y_{1,k} = S_{n_d} X_{1,k} \oplus B_{1,k} S_{n_c} X_{2,k}, \tag{3}$$
$$Y_{2,k} = S_{n_d} X_{2,k} \oplus B_{2,k} S_{n_c} X_{1,k}. \tag{4}$$

Let $q \triangleq \max\{n_d, n_c\}$. In (3) and (4), $\boldsymbol{X}_{i,k} \in \mathbb{F}_2^{q \times T}$ and $\boldsymbol{Y}_{i,k} \in \mathbb{F}_2^{q \times T}$, $i = 1, 2$. The interference states $B_{i,k}$, $i = 1, 2$, $k = 1, \dots, K$, are sequences of i.i.d. Bernoulli random variables with activation probability p.

Regarding the sequences B_1^K and B_2^K, we consider two cases: (i) B_1^K and B_2^K are independent of each other and (ii) B_1^K and B_2^K are fully correlated sequences, i.e., $B_1^K = B_2^K$. For both cases we assume that the sequences are independent of the messages W_1 and W_2.

We shall define the normalized interference level as $\alpha \triangleq \frac{n_c}{n_d}$, based on which we can divide the interference into the following regions (a similar division was used by Jafar and Vishwanath [20]):

- *very weak interference (VWI)* for $\alpha \le \frac{1}{2}$,
- *weak interference (WI)* for $\frac{1}{2} < \alpha \le \frac{2}{3}$,
- *moderate interference (MI)* for $\frac{2}{3} < \alpha \le 1$,
- *strong interference (SI)* for $1 < \alpha \le 2$,
- *very strong interference (VSI)* for $2 < \alpha$.

2.1. Quasi-Static Channel

The channel defined in (3) and (4) may experience a slowly-varying change on the interference state. In this case, the duration of each of the transmitted codewords of length $N = KT$ is smaller than the coherence time T of the channel and the interference state stays constant over the duration of each codeword, i.e., $K = 1$, $T = N$. In the wireless communications literature such a channel is usually referred to as a quasi-static channel [21] (Section 5.4.1). In this scenario, the rate pair of achievable rates (R_1, R_2) is dominated by the worst case, which corresponds to the presence of interference at both receivers. However, in absence of interference, it is possible to communicate at a higher date rate, so planning a system for the worst case may be too pessimistic. Assuming that the receivers have access to the interference states, the transmitters could send opportunistic messages that are decoded only if the interference is absent, in addition to the regular messages that are decoded irrespective of the interference state. We make the notion of opportunistic messages and rates precise in the subsequent paragraphs.

Let $U_{i,k}$ indicate the level of CSI available at the transmitter side in coherence block k, and let $V_{i,k}$ indicate the level of CSI at the receiver side in coherence block k:

1. local CSIR:　　$U_{i,k} = \varnothing$ and $V_{i,k} = B_{i,k}$, $i = 1, 2$, $k = 1, \dots, K$,
2. local CSIRT:　$U_{i,k} = V_{i,k} = B_{i,k}$, $i = 1, 2$, $k = 1, \dots, K$,
3. global CSIRT:　$U_{i,k} = V_{i,k} = (B_{1,k}, B_{2,k})$, $i = 1, 2$, $k = 1, \dots, K$.

We define the set of opportunistic messages according to the level of CSI at the receiver as $\{\Delta W_i(\cdot)\} \triangleq \{\Delta W_i(v_i), v_i \in \mathcal{V}_i\}$, where $\mathcal{V}_i$ denotes the set of possible interference states V_i. Specifically,

1. for local CSIR: $\{\Delta W_i(\cdot)\} = \{\Delta W_i(1), \Delta W_i(0)\}$, $i = 1, 2$,
2. for local CSIRT: $\{\Delta W_i(\cdot)\} = \{\Delta W_i(1), \Delta W_i(0)\}$, $i = 1, 2$,
3. for global CSIRT: $\{\Delta W_i(\cdot)\} = \{\Delta W_i(00), \Delta W_i(01), \Delta W_i(10), \Delta W_i(11)\}$, $i = 1, 2$.

Then, we define an opportunistic code as follows.

Definition 1 (Opportunistic code for the bursty IC)**.** *An $(N, R_1, R_2, \{\Delta R_1(\cdot)\}, \{\Delta R_2(\cdot)\})$ opportunistic code for the bursty IC is defined as:*

1. *two independent messages W_1 and W_2 uniformly distributed over the message sets $\mathcal{W}_i \triangleq \{1, 2, \dots, 2^{NR_i}\}$, $i = 1, 2$;*
2. *two independent sets of opportunistic messages $\{\Delta W_1(\cdot)\}$ and $\{\Delta W_2(\cdot)\}$ uniformly distributed over the message sets $\Delta \mathcal{W}_i(v_i) \triangleq \{1, 2, \dots, 2^{N\Delta R_i(v_i)}\}$, $v_i \in \mathcal{V}_i$, $i = 1, 2$,*
3. *two encoders: $f_i : (W_i, \{\Delta W_i(\cdot)\}, U_i) \mapsto \boldsymbol{X}_i$, $i = 1, 2$,*

4.   two decoders: $g_i : (\mathbf{Y}_i, V_i) \mapsto (\hat{W}_i, \Delta \hat{W}_i(V_i))$, $i = 1, 2$.

Here $\hat{W}_i$ and $\Delta \hat{W}_i(V_i)$ denote the decoded message and the decoded opportunistic message, respectively. We set $\Delta R_i(1) = 0$, $i = 1, 2$ *(for local CSIR/CSIRT) and $\Delta R_i(11) = 0$ (for global CSIRT).*

To better distinguish the rates (R_1, R_2) from the opportunistic rates $\{\Delta R_i(\cdot)\}$, $i = 1, 2$, we shall refer to (R_1, R_2) as worst-case rates, because the corresponding messages can be decoded even if the channel is in its worst state (see also Definition 2).

Definition 2 (Achievable opportunistic rates). *A rate tuple $\big(R_1, R_2, \{\Delta R_1(\cdot)\}, \{\Delta R_2(\cdot)\}\big)$ is achievable if there exists a sequence of codes $\big(N, R_1, R_2, \{\Delta R_1(\cdot)\}, \{\Delta R_2(\cdot)\}\big)$ such that*

$$lCl \Pr\{\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2\} \to 0 \quad as \quad N \to \infty \tag{5}$$

and

$$lCl \Pr\big\{(\hat{W}_1, \Delta \hat{W}_1(V_1)) \neq (W_1, \Delta W_1(V_1)) | V_1 = v_1\big\} \to 0 \quad as \quad N \to \infty, \ v_1 \in \mathcal{V}_1, \tag{6}$$

$$lCl \Pr\big\{(\hat{W}_2, \Delta \hat{W}_2(V_2)) \neq (W_2, \Delta W_2(V_2)) | V_2 = v_2\big\} \to 0 \quad as \quad N \to \infty, \ v_2 \in \mathcal{V}_2. \tag{7}$$

The capacity region is the closure of the set of achievable rate tuples [1](Sec. 6.1). We define the worst-case sum rate as $R \triangleq R_1 + R_2$ and the opportunistic sum rate as $\Delta R(V_1, V_2) \triangleq \Delta R_1(V_1) + \Delta R_2(V_2)$. The worst-case sum capacity C is the supremum of all achievable worst-case sum rates, the opportunistic sum capacity $\Delta C(V_1, V_2)$ is the supremum of all opportunistic sum rates, and the total sum capacity is defined as $C + \Delta C(V_1, V_2)$. Note that the opportunistic sum capacity depends on the worst-case sum rate.

Remark 1. *The worst-case sum rate and opportunistic sum rates in the quasi-static setting depend only on the collection of possible interference states: for independent interference states we have $\mathbf{B} \in \{00, 01, 10, 11\}$, and for fully correlated interference states we have $\mathbf{B} \in \{00, 11\}$. In principle, our proof techniques could also be applied to analyze other collections of interference states.*

Remark 2. *In the CSIRT setting the transmitters have access to the interference state. Therefore, in this setting the messages are strictly speaking not opportunistic. Instead, transmitters can adapt their rate based on the state of the interference links, which is sometimes referred to as rate adaptation in the literature.*

2.2. Ergodic Channel

In this setup, we shall restrict ourselves to codes whose blocklength N is an integer multiple of the coherence time T. A codeword of length $N = KT$ thus spans K independent channel realizations.

Definition 3 (Code for the bursty IC). *A $\big(K, T, R_1, R_2\big)$ code for the bursty IC is defined as:*

1.   two independent messages W_1 and W_2 uniformly distributed over the message sets $\mathcal{W}_i \triangleq \{1, 2, \ldots, 2^{KTR_i}\}$, $i = 1, 2$;
2.   two encoders: $f_i : (W_i, U_i^K) \mapsto \mathbf{X}_i^K$, $i = 1, 2$;
3.   two decoders: $g_i : (Y_i^K, V_i^K) \mapsto \hat{W}_i$, $i = 1, 2$.

Here $\hat{W}_i$ denotes the decoded message, and U_i^K and V_i^K indicate the level of CSI at the transmitter and receiver side, respectively, which are defined as for the quasi-static channel in Section 2.1.

Definition 4 (Ergodic achievable rates). *A rate pair (R_1, R_2) is achievable for a fixed T if there exists a sequence of codes (K, T, R_1, R_2) (parametrized by K) such that*

$$\Pr\{\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2\} \to 0 \quad as \quad K \to \infty. \tag{8}$$

The capacity region is the closure of the set of achievable rate pairs. We define the sum rate as $R \triangleq R_1 + R_2$, the sum capacity C is the supremum of all achievable sum rates.

2.3. The Sum Capacities of the Non-Bursty and the Quasi-Static Bursty IC

When the activation probability p is 1, we recover in both the ergodic and quasi-static scenarios the deterministic IC. For a general deterministic IC the capacity region was obtained in [16] (Th. 1) and then by Bresler and Tse in [17] for a specific deterministic IC. For completeness, we present the sum capacity region for the deterministic non-bursty IC in the following theorem.

Theorem 1. *The sum capacity region of the two-user deterministic IC is equal to the union of the set of all sum rates R satisfying*

$$R \leq 2n_d \tag{9}$$

$$R \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{10}$$

$$R \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{11}$$

Proof. The proof is given in [16] (Section II). For the achievability bounds, El Gamal and Costa [16] (Theorem 1) use the Han-Kobayashi scheme [22] for a general IC. Bresler and Tse [17] (Section 4) use a specific Han-Kobayashi strategy for the special case of the LDM. Jafar and Vishwanath [20] present an alternative achievability scheme for the K-user IC, which particularized for the two-user IC will be referenced in this work. □

We can achieve the sum rates (9) and (11) over the quasi-static channel by treating the bursty IC as a non-bursty IC. The following theorem demonstrates that this is the largest achievable worst-case sum rate irrespective of the availability of CSI and the correlation between B_1 and B_2.

Theorem 2 (Sum capacity for the quasi-static bursty IC). *For $0 \leq p \leq 1$, the worst-case sum capacity of the bursty IC is equal to the supremum of the set of sum rates R satisfying*

- *For $p = 0$,*

$$R \leq 2n_d. \tag{12}$$

- *For $0 < p \leq 1$*

$$R \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{13}$$

$$R \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{14}$$

Proof. The converse bounds are proved in Appendix A.1. Achievability follows directly from Theorem 1 by treating the bursty IC as a non-bursty IC. □

Theorem 2 shows that the worst-case sum capacity does not depend on the level of CSI available at the transmitter and receiver side. However, this is not the case for the opportunistic rates as we will see in the next sections.

Remark 3. *In principle, one could reduce the worst-case rates in order to increase the opportunistic rates. However, it turns out that such a strategy is not beneficial in terms of total rates $R_i + \Delta R_i(V_i)$, $i = 1, 2$. In other words, setting $\Delta R_i(1) = 0$, $i = 1, 2$ (for local CSIR/CSIRT) and $\Delta R_i(11) = 0$ (for global CSIRT), as we have done in Definition 2, incurs no loss in total rate. Furthermore, in most cases it is preferable to maximize the worst-case rate, since it can be guaranteed irrespective of the interference state.*

3. Local CSIR

For the quasi-static and ergodic setups, described in Sections 2.1 and 2.2, respectively, we derive converse and achievability bounds for the independent and fully correlated scenarios when the interference state is only available at the receiver side.

3.1. Quasi-Static Channel

3.1.1. Independent Case

We present converse and achievability bounds for local CSIR when B_1 and B_2 are independent. The converse bounds are derived for local CSIRT, hence they also apply to this case. Since converse and achievability bounds coincide, this implies that local CSI at the transmitter is not beneficial in the quasi-static setup.

Theorem 3 (Opportunistic sum capacity for local CSIR/CSIRT). *Assume that B_1 and B_2 are independent of each other. For $0 < p < 1$, the opportunistic sum capacity region is the union of the set of rate tuples $(R, \{\Delta R_1(b_1) + \Delta R_2(b_2), b_i \in \{0, 1\}\})$, where $\Delta R_1(1) = \Delta R_2(1) = 0$, and R, $\Delta R_1(0)$ and $\Delta R_2(0)$ satisfy* (12)–(14) *and*

$$R + \Delta R_1(0) + \Delta R_2(0) \le 2n_d \tag{15}$$

$$R + \Delta R_1(0) \le (n_d - n_c)^+ + \max(n_d, n_c) \tag{16}$$

$$R + \Delta R_2(0) \le (n_d - n_c)^+ + \max(n_d, n_c). \tag{17}$$

Proof. The converse bounds are proved in Appendix A.2 and the achievability bounds are proved in Appendix A.3. □

Remark 4. *The converse bounds in Theorem 3 coincide with those in [3] (Theorem 2.1), particularized for the symmetric setting. Theorem 3, however, is proven for local CSIRT, which is not considered in the model from [3]. The proof included in Appendix A.2 is based on an information density approach and provides a unified framework for treating local CSIR, local CSIRT and global CSIRT, as will be shown in Section 5.*

As discussed in Remark 3, one could reduce the worst-case sum rate R and increase the opportunistic rates $\Delta R(V_1, V_2)$. However, in the case of one-shot transmission this is not desirable, since the worst-case sum rate is the only rate that can be guaranteed irrespective of the interference state. (With one-shot transmission we refer to the case where we transmit one codeword of length N over the quasi-static channel. This is in contrast to the case discussed, e.g., in Section 3.3, where we are interested in transmitting many codewords, each over N channel uses of independent quasi-static channels.) Thus, one is typically interested in the opportunistic sum capacity when the worst-case rate R is maximized. For this case, the results of Theorem 3 are summarized in Table 1 for the VWI, WI, MI and SI regions.

Table 1. Opportunistic sum capacity for local CSIR when the worst-case sum rate is maximized.

Rates	VWI	WI	MI	SI
C	$2(n_d - n_c)$	$2n_c$	$2n_d - n_c$	n_c
$\Delta C(00)$	$2n_c$	$2(2n_d - 3n_c)$	0	0
$\Delta C(01)/\Delta C(10)$	n_c	$2n_d - 3n_c$	0	0

Observe that converse and achievability bounds coincide. Further observe that opportunistic messages can only be transmitted reliably for VWI or WI. In the other interference regions, the opportunistic sum capacity is zero.

3.1.2. Fully Correlated Case

Assume now that the sequences B_1 and B_2 are fully correlated ($B_1 = B_2$). For local CSIR, the correlation between B_1 and B_2 has no influence on the opportunistic sum capacity region. Indeed, in this case the channel inputs are independent of (B_1, B_2) and the opportunistic sum capacity region

of the quasi-static bursty IC depends on (B_1, B_2) only via the marginal distributions of B_i, $i = 1, 2$. Hence, it follows that Theorem 3 as well as Table 1 apply also to the fully correlated case and local CSIR scenario. For completeness, a proof of the converse part is given in Appendix A.4. The achievability part is included in Appendix A.3.

3.2. Ergodic Channel

3.2.1. Independent Case

For the case where the sequences B_1^K and B_2^K are independent of each other, we have the following theorems.

Theorem 4 (Converse bounds for local CSIR). *Assume that B_1^K and B_2^K are independent of each other. The sum rate R for the bursty IC is upper-bounded by*

$$R \leq 2\frac{1-p}{1+p}n_d + 2\frac{p}{1+p}\left[(n_d - n_c)^+ + \max(n_d, n_c)\right] \tag{18}$$

and

$$R \leq \begin{cases} 2(1-2p)n_d + 2p\left[(n_d - n_c)^+ + \max(n_d, n_c)\right] & p \leq \frac{1}{2}, \\ 2(1-p)\left[(n_d - n_c)^+ + \max(n_d, n_c)\right] + 2(2p-1)\left[\max\{(n_d - n_c)^+, n_c\}\right] & p > \frac{1}{2}. \end{cases} \tag{19}$$

Proof. Bound (18) coincides with [7] (Equation (3)). Specifically, [7] (Equation (3)) derives (18) for the considered channel model with $T = 1$ and feedback. The proof for this bound under local CSIRT (without feedback) is given in Appendix B.1. Bound (19) coincides with [23] (Lemma A.1). Specifically, [23] (Lemma A.1) derives (19) for the model considered with $T = 1$. The proof of [23] (Lemma A.1) directly generalizes to arbitrary T. □

Theorem 5 (Achievability bounds for local CSIR). *Assume that B_1^K and B_2^K are independent of each other. The following sum rate R is achievable over the bursty IC:*

$$R = \begin{cases} 2(1-2p)n_d + 2p\left[(n_d - n_c)^+ + \max(n_d, n_c)\right], & p \leq \frac{1}{2}, \\ \min\left\{(n_d - n_c)^+ + \max(n_d, n_c), \right. \\ \left. 2(1-p)\left[(n_d - n_c)^+ + \max(n_d, n_c)\right] + 2(2p-1)\left[\max\{(n_d - n_c)^+, n_c\}\right]\right\}, & p > \frac{1}{2}. \end{cases} \tag{20}$$

Proof. The achievability scheme for VWI for all values of p, and for WI and MI when $0 \leq p \leq \frac{1}{2}$, is described in Appendix B.2.1. The achievability scheme for WI and $\frac{1}{2} < p \leq 1$ is described in Appendix B.2.2. The scheme for SI and $0 \leq p \leq \frac{1}{2}$ is summarized in Appendix B.2.3. For MI and SI when $\frac{1}{2} < p \leq 1$, the achievability bound in the theorem corresponds to the one of the non-bursty IC [20]. This also implies that in this sub-region we do not exploit the burstiness of the IC. □

Remark 5. *The achievability schemes presented in Theorem 5 are similar to those described in [11,14]. They achieve the capacity region by applying point-to-point erasure codes with appropriate rates at each transmitter and using either treating-interference-as-erasure or interference-decoding at each receiver. Specifically, we apply treating-interference-as-erasure in the VWI region and for all values of p, and for all interference regions, except VSI, and $p \leq \frac{1}{2}$. Interference-decoding at each receiver is applied in the MI and SI regions for $p > \frac{1}{2}$.*

Remark 6. *Wang et al. claim in [23] (Lemma A.1) that the converse bound (18) is tight for $0 \leq p \leq \frac{1}{2}$ without providing an achievability bound. Instead, they refer to Khude et al. [3] for the inner bound which, alas, does not apply to the ergodic setup. While it is possible to adapt the achievability schemes considered in [3] to prove (20),*

a number of steps are required. For completeness, we include the achievability schemes for the ergodic setup and
$0 \le p \le \frac{1}{2}$ *in Appendix B.2.1.*

Table 2 summarizes the results of Theorems 4 and 5. We write the sum capacities in bold face
when the converse and achievability bounds match. In Table 2, we define

$$\mathbf{C}_{\mathrm{LMI}} \triangleq \min \left\{ 2[2(n_d - n_c) + p(3n_c - 2n_d)], 2 \left[\frac{1-p}{1+p} n_d + \frac{p}{1+p}(2n_d - n_c) \right] \right\} \tag{21}$$

$$\mathbf{C}_{\mathrm{LSI}} \triangleq \min \left\{ 2pn_c, 2 \left[\frac{1-p}{1+p} n_d + \frac{p}{1+p} n_c \right] \right\} \tag{22}$$

where "L" stands for "local CSIR".

Table 2. Sum capacity for local CSIR.

Regions	$p \le \frac{1}{2}$	$p > \frac{1}{2}$
VWI	$\mathbf{2(n_d - pn_c)}$	$\mathbf{2(n_d - pn_c)}$
WI	$\mathbf{2(n_d - pn_c)}$	$\mathbf{4(n_d - n_c) + 2p(3n_c - 2n_d)}$
MI	$\mathbf{2(n_d - pn_c)}$	$2n_d - n_c \le R \le \mathbf{C}_{\mathrm{LMI}}$
SI	$\mathbf{2(1 - 2p)n_d + 2pn_c}$	$n_c \le R \le \mathbf{C}_{\mathrm{LSI}}$

3.2.2. Fully Correlated Case

For local CSIR, the dependence between B_1^K and B_2^K has no influence on the capacity region.
Indeed, in this case the channel inputs are independent of (B_1^K, B_2^K) and decoder i has only access to $B_{i,k}$
and $(S_{n_d} X_{i,k} \oplus B_{i,k} S_{n_c} X_{j,k})$, $k = 1, \ldots, K$, $j = 3 - i$ and $i = 1, 2$. Furthermore, $\Pr\{\hat{W}_1 \ne W_1 \cup \hat{W}_2 \ne W_2\}$
vanishes as $K \to \infty$ if, and only if, $\Pr\{\hat{W}_i \ne W_i\}$, $i = 1, 2$, vanishes as $K \to \infty$. Since $\Pr(\hat{W}_i \ne W_i)$
depends only on B_i^K, the capacity region of the bursty IC depends on (B_1^K, B_2^K) only via the marginal
distributions of B_1^K and B_2^K. Hence, Theorems 4 and 5 as well as Table 2 apply also to the case where
$B_1^K = B_2^K$. This is consistent with the observation by Sato [24] that "the capacity region is the same for
all two-user channels that have the same marginal probabilities".

3.3. Quasi-Static vs. Ergodic Setup

In general, the sum capacities of the quasi-static and ergodic channels cannot be compared,
because in the former case we have a set of sum capacities (worst case and opportunistic), whereas in
the latter case only one is defined. To allow for a comparison, we introduce for the quasi-static channel
the average sum capacity as

$$\bar{C} \triangleq \sup_{(R, \Delta R_1(0), \Delta R_2(0))} \{ R + (1 - p)(\Delta R_1(0) + \Delta R_2(0)) \} \tag{23}$$

where the suprema is over all tuples $(R, \Delta R_1(0), \Delta R_2(0))$ that satisfy (12)–(17). Intuitively, the average
rate corresponds to the case where we send many messages over independent quasi-static fading
channels. By the law of large numbers, a fraction of p transmissions will be affected by interference,
the remaining transmissions will be interference-free. Table 3 summarizes the average sum capacity
for the different interference regions.

By comparing Tables 2 and 3, we can observe that for $p \le \frac{1}{2}$ and all interference regions, and
for $p > \frac{1}{2}$ and VWI/WI, the average sum capacity in the quasi-static setup coincides with the sum
capacity in the ergodic setup. For $p > \frac{1}{2}$, and MI/SI (where converse and achievability bounds do not
coincide), the average sum capacities in the quasi-static setup coincide with the achievability bounds
of the ergodic setup.

Table 3. Average sum capacities for local CSIR.

Regions	$p \leq \frac{1}{2}$	$p > \frac{1}{2}$
VWI	$2(n_d - pn_c)$	$2(n_d - pn_c)$
WI	$2(n_d - pn_c)$	$4(n_d - n_c) + 2p(3n_c - 2n_d)$
MI	$2(n_d - pn_c)$	$2n_d - n_c$
SI	$2(1 - 2p)n_d + 2pn_c$	n_c

4. Local CSIRT

For the quasi-static and ergodic setups, we present converse and achievability bounds when transmitters and receivers have access to their corresponding interference states. We shall only consider the independent case here, because when $B_1^K = B_2^K$ local CSIRT coincides with global CSIRT, which will be discussed in Section 5.

4.1. Quasi-Static Channel

For the quasi-static channel, the converse and achievability bounds were already presented in Theorem 3 in Section 3.1.1. Indeed, the converse bounds were derived for local CSIRT, whereas the achievability bounds in that theorem were derived for local CSIR. Since these bounds coincide for all interference regions and all probabilities of $0 < p < 1$ it follows that, for the quasi-static channel, availability of local CSI at the transmitter in addition to local CSI at the receiver is not beneficial. The converse and achievability bounds are then given in Theorem 3.

4.2. Ergodic Channel

The converse bound (18) presented in Theorem 4 was derived for local CSIRT, so it applies to the case at hand. We next present achievability bounds for this setup that improve upon those for CSIR. The aim of these bounds is to provide computable expressions showing that local CSIRT outperforms local CSIR in the whole range of the α parameter. While the particular achievability schemes are sometimes involved, the intuition behind these schemes can be explained with the following toy example.

Example 1. *Let us assume that $n_d = n_c = T = 1$, and suppose that at time k the transmitters send the bits $(B_{1,k}, B_{2,k}) \in \{0,1\}^2$. If there is no interference, then receiver i receives $X_{i,k}$. If there is interference, then receiver i receives $X_{1,k} \oplus X_{2,k}$. Consequently, the channel flips $X_{1,k}$ if $B_{1,k} = X_{2,k} = 1$, and it flips $X_{2,k}$ if $B_{2,k} = X_{1,k} = 1$. It follows that each transmitter-receiver pair experiences a binary symmetric channel (BSC) with a given crossover probability that depends on p and on the probabilities that (X_1, X_2) are one. Specifically, let*

$$P_{X_1|B_1}(X_1 = 1 | B_1 = 0) \triangleq p_1 \tag{24}$$

$$P_{X_1|B_1}(X_1 = 1 | B_1 = 1) \triangleq p_2 \tag{25}$$

$$P_{X_2|B_2}(X_2 = 1 | B_2 = 0) \triangleq q_1 \tag{26}$$

$$P_{X_2|B_2}(X_2 = 1 | B_2 = 1) \triangleq q_2 \tag{27}$$

and define $p_3 \triangleq (1 - p)p_1 + pp_2$ and $q_3 \triangleq (1 - p)q_1 + pq_2$, which are the crossover probabilities of the BSCs experienced by receivers 1 and 2, respectively, when they are affected by interference. By drawing for each user two codebooks (one for $B_{i,k} = 0$ and one for $B_{i,k} = 1$) i.i.d. at random according to the probabilities p_1, p_2, q_1, and q_2, and by following a random-coding argument, it can be shown that this scheme achieves the sum rate

$$lClR = (1 - p)[H_b(p_1) + H_b(q_1)] + p[H_{sum}(p_2, q_3) - H_b(q_3)] + p[H_{sum}(q_2, p_3) - H_b(p_3)]. \tag{28}$$

This expression holds for any set of parameters (p_1, p_2, q_1, q_2), and the largest sum rate achieved by this scheme is obtained by maximizing over $(p_1, p_2, q_1, q_2) \in [0, \frac{1}{2}]^4$.

In the following, we present the achievable sum rates that can be obtained by generalizing the above achievability scheme to general n_d and n_c. The achievability schemes that achieve these rates are presented in Appendix D. The largest achievable sum rates can then be obtained by numerically maximizing over the parameters $(p_1, p_2, q_1, q_2, \ldots)$ (which depend on the interference region).

1. For the VWI region, we achieve the sum rate

$$R = 2(n_d - pn_c). \tag{29}$$

2. For the WI region, we can achieve for any $(p_1, p_2, q_1, q_2) \in [0, \frac{1}{2}]^4$

$$R_1 = (n_d - n_c) + (1 - p)[(n_d - n_c) + (2n_c - n_d)H_b(p_1)] + p(2n_c - n_d)(1 - H_b(q_3)) \tag{30}$$

$$R_2 = (n_d - n_c) + (1 - p)[(n_d - n_c) + (2n_c - n_d)H_b(q_1)] + p(2n_c - n_d)(1 - H_b(p_3)) \tag{31}$$

where $p_3 = (1 - p)p_1 + pp_2$ and $q_3 = (1 - p)q_1 + pq_2$.

3. To present the achievable rates for MI, we need to divide the region into the following four subregions:

    (a) For $\frac{2}{3} \le \alpha \le \frac{3}{4}$, we can achieve for any $(p_1, p_2, \tilde{p}_1, \tilde{p}_2, \hat{p}_1, q_1, q_2, \tilde{q}_1, \tilde{q}_2, \hat{q}_1) \in [0, \frac{1}{2}]^{10}$ and $(\eta_1, \gamma_1) \in [\frac{1}{2}, 1]^2$

$$
\begin{aligned}
R_1 =&(n_d - n_c) \\
&+ (1 - p)\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_b(\eta_1) + H_b(\hat{p}_1) + H_b(p_1)) + \left(\tfrac{4n_d - 5n_c}{2}\right)H_b(\tilde{p}_1) + (n_d - n_c)\right] \\
&+ p\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(1 + H_{\text{sum}}(p_2, \tilde{\gamma}) - H_b(\tilde{\gamma}) + H_{\text{sum}}(\tilde{p}_2, q_3) - H_b(q_3) - H_b(\hat{q}_3)) \right. \\
&\left. + \left(\tfrac{4n_d - 5n_c}{2}\right)(1 - H_b(\tilde{q}_3))\right]
\end{aligned}
\tag{32}
$$

where $q_3 = (1 - p)q_1 + pq_2$, $\tilde{q}_3 = (1 - p)\tilde{q}_1 + p\tilde{q}_2$, $\hat{q}_3 = (1 - p)\hat{q}_1$, and $\tilde{\gamma} = p + \gamma_1(1 - p)$, and

$$
\begin{aligned}
R_2 =&(n_d - n_c) \\
&+ (1 - p)\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_b(\gamma_1) + H_b(\hat{q}_1) + H_b(q_1)) + \left(\tfrac{4n_d - 5n_c}{2}\right)H_b(\tilde{q}_1) + (n_d - n_c)\right] \\
&+ p\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(1 + H_{\text{sum}}(q_2, \tilde{\eta}) - H_b(\tilde{\eta}) + H_{\text{sum}}(\tilde{q}_2, p_3) - H_b(p_3) - H_b(\hat{p}_3)) \right. \\
&\left. + \left(\tfrac{4n_d - 5n_c}{2}\right)(1 - H_b(\tilde{p}_3))\right]
\end{aligned}
\tag{33}
$$

where $p_3 = (1 - p)p_1 + pp_2$, $\tilde{p}_3 = (1 - p)\tilde{p}_1 + p\tilde{p}_2$, $\hat{p}_3 = (1 - p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1 - p)$.

Remark 7. *After combining (32) and (33), η_1 and γ_1 appear only through the functions $H_b(\eta_1) - H_b(p + \eta_1(1 - p))$ and $H_b(\gamma_1) - H_b(p + \gamma_1(1 - p))$, respectively. Hence, η_1 and γ_1 can be optimized separately from the remaining terms.*

    (b) For $\frac{3}{4} \le \alpha \le \frac{4}{5}$, we can achieve for any $(p_1, p_2, \tilde{p}_1, \tilde{p}_2, \hat{p}_1, q_1, q_2, \tilde{q}_1, \tilde{q}_2, \hat{q}_1) \in [0, \frac{1}{2}]^{10}$ and $(\eta_1, \gamma_1) \in [\frac{1}{2}, 1]^2$

$$R_1 = (n_d - n_c)$$
$$+ (1-p)\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_b(p_1) + H_b(\eta_1) + H_b(\hat{p}_1)) + (\tfrac{4n_d - 5n_c}{2})H_b(\tilde{p}_1) + (n_d - n_c)\right]$$
$$+ p\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_{\text{sum}}(p_2, \tilde{\gamma}) - H_b(\tilde{\gamma}) + 1 - H_b(\hat{q}_3))\right.$$
$$+ \left.\left(\tfrac{4n_d - 5n_c}{2}\right)(H_{\text{sum}}(\tilde{p}_2, q_3) - H_b(q_3) + 1 - H_b(\tilde{q}_3))\right] \tag{34}$$

where $q_3 = (1-p)q_1 + pq_2$, $\tilde{q}_3 = (1-p)\tilde{q}_1 + p\tilde{q}_2$, $\hat{q}_3 = (1-p)\hat{q}_1$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$R_2 = (n_d - n_c)$$
$$+ (1-p)\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_b(q_1) + H_b(\gamma_1) + H_b(\hat{q}_1)) + (\tfrac{4n_d - 5n_c}{2})H_b(\tilde{q}_1) + (n_d - n_c)\right]$$
$$+ p\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_{\text{sum}}(q_2, \tilde{\eta}) - H_b(\tilde{\eta}) + 1 - H_b(\hat{p}_3))\right.$$
$$+ \left.\left(\tfrac{4n_d - 5n_c}{2}\right)(H_{\text{sum}}(\tilde{q}_2, p_3) - H_b(p_3) + 1 - H_b(\tilde{p}_3))\right] \tag{35}$$

where $p_3 = (1-p)p_1 + pp_2$, $\tilde{p}_3 = (1-p)\tilde{p}_1 + p\tilde{p}_2$, $\hat{p}_3 = (1-p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1-p)$. Remark 7 also applies to the parameters η_1 and γ_1 in (34) and (35).

(c) For $\tfrac{4}{5} \le \alpha \le \tfrac{6}{7}$, we can achieve for any $(p_1, p_2, \hat{p}_1, q_1, q_2, \hat{q}_1) \in \left[0, \tfrac{1}{2}\right]^6$ and $(\eta_1, \eta', \gamma_1, \gamma') \in \left[\tfrac{1}{2}, 1\right]^4$

$$R_1 = (n_d - n_c)$$
$$+ (1-p)\left[\left(\tfrac{5n_c - 4n_d}{2}\right)(1 + H_b(\eta')) + (n_d - n_c)(1 + H_b(p_1) + H_b(\eta_1) + H_b(\hat{p}_1))\right]$$
$$+ p\left[\left(\tfrac{5n_c - 4n_d}{2}\right)(1 - H_b(\tilde{\gamma}) + H_{\text{sum}}(p_2, \gamma') - H_b(\gamma')\right.$$
$$+ H_{\text{sum}}(\eta'(1 - \tilde{\gamma}) + (1 - \eta')\tilde{\gamma}, q_3) - H_b(q_3))$$
$$+ \left.\left(\tfrac{6n_d - 7n_c}{2}\right)(H_{\text{sum}}(p_2, \tilde{\gamma}) - H_b(\tilde{\gamma})) + (n_d - n_c)(1 - H_b(\hat{q}_3))\right] \tag{36}$$

where $q_3 = (1-p)q_1 + pq_2$, $\hat{q}_3 = (1-p)\hat{q}_1$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$R_2 = (n_d - n_c)$$
$$+ (1-p)\left[\left(\tfrac{5n_c - 4n_d}{2}\right)(1 + H_b(\gamma')) + (n_d - n_c)(1 + H_b(q_1) + H_b(\gamma_1) + H_b(\hat{q}_1))\right]$$
$$+ p\left[\left(\tfrac{5n_c - 4n_d}{2}\right)(1 - H_b(\tilde{\eta}) + H_{\text{sum}}(q_2, \eta') - H_b(\eta')\right.$$
$$+ H_{\text{sum}}(\gamma'(1 - \tilde{\eta}) + (1 - \gamma')\tilde{\eta}, p_3) - H_b(p_3))$$
$$+ \left.\left(\tfrac{6n_d - 7n_c}{2}\right)(H_{\text{sum}}(q_2, \tilde{\eta}) - H_b(\tilde{\eta})) + (n_d - n_c)(1 - H_b(\hat{p}_3))\right] \tag{37}$$

where $p_3 = (1-p)p_1 + pp_2$, $\hat{p}_3 = (1-p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1-p)$.

(d) For $\tfrac{6}{7} \le \alpha \le 1$ we can achieve for any $(p_1, p_2, \hat{p}_1, q_1, q_2, \hat{q}_1) \in \left[0, \tfrac{1}{2}\right]^6$ and $(\eta_1, \eta', \gamma_1, \gamma') \in \left[\tfrac{1}{2}, 1\right]^4$

$$R_1 = (n_d - n_c)$$
$$+ (1-p)\left[(6n_c - 5n_d)H_b(p_1) + (n_d - n_c)(2 + H_b(\eta_1) + H_b(\eta') + H_b(\hat{p}_1))\right]$$
$$+ p\left[(n_d - n_c)(2 - H_b(\tilde{\gamma}) - H_b(\hat{q}_3) + H_{\text{sum}}(\eta'(1 - \tilde{\gamma}) + (1 - \eta')\tilde{\gamma}, q_3) - H_b(q_3))\right.$$
$$+ (n_d - n_c)(H_{\text{sum}}(p_2, \gamma') - H_b(\gamma'))$$
$$+ \left.(7n_c - 6n_d)(H_{\text{sum}}(p_2, q_3) - H_b(q_3))\right] \tag{38}$$

where $q_3 = (1-p)q_1 + pq_2$, $\hat{q}_3 = (1-p)\hat{q}_1$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$
\begin{aligned}
R_2 =& (n_d - n_c) \\
&+ (1-p)\big[(6n_c - 5n_d)H_b(q_1) + (n_d - n_c)\big(2 + H_b(\gamma_1) + H_b(\gamma') + H_b(\hat{q}_1)\big)\big] \\
&+ p\big[(n_d - n_c)\big(2 - H_b(\tilde{\eta}) - H_b(\hat{p}_3) + H_{sum}(\gamma'(1-\tilde{\eta}) + (1-\gamma')\tilde{\eta}, p_3) - H_b(p_3)\big) \\
&+ (n_d - n_c)\big(H_{sum}(q_2, \eta') - H_b(\eta')\big) \\
&+ (7n_c - 6n_d)\big(H_{sum}(q_2, p_3) - H_b(p_3)\big)\big]
\end{aligned}
\tag{39}
$$

where $p_3 = (1-p)p_1 + pp_2$, $\hat{p}_3 = (1-p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1-p)$.

4. To present the achievable rates for SI, we divide the region into the following four subregions:

(a) For $1 \leq \alpha \leq \frac{6}{5}$, we can achieve for any $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$ and $(\eta_1, \eta', \gamma_1, \gamma') \in \left[\frac{1}{2}, 1\right]^4$

$$
\begin{aligned}
R_1 =& (n_c - n_d) + (1-p)\big[(5n_d - 4n_c)H_b(p_1) + (n_c - n_d)\big(1 + H_b(\eta_1) + H_b(\eta')\big)\big] \\
&+ p\big[(n_c - n_d)\big(1 - H_b(\tilde{\gamma}) + H_{sum}(\eta'(1-\tilde{\gamma}) + (1-\eta')\tilde{\gamma}, q_3) - H_b(q_3)\big) \\
&+ (n_c - n_d)\big(H_{sum}(p_2, \gamma') - H_b(\gamma')\big) \\
&+ (6n_d - 5n_c)\big(H_{sum}(p_2, q_3) - H_b(q_3)\big)\big]
\end{aligned}
\tag{40}
$$

where $q_3 = (1-p)q_1 + pq_2$ and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$
\begin{aligned}
R_2 =& (n_c - n_d) + (1-p)\big[(5n_d - 4n_c)H_b(q_1) + (n_c - n_d)\big(1 + H_b(\gamma_1) + H_b(\gamma')\big)\big] \\
&+ p\big[(n_c - n_d)\big(1 - H_b(\tilde{\eta}) + H_{sum}(\gamma'(1-\tilde{\eta}) + (1-\gamma')\tilde{\eta}, q_3) - H_b(p_3)\big) \\
&+ (n_c - n_d)\big(H_{sum}(q_2, \eta') - H_b(\eta')\big) \\
&+ (6n_d - 5n_c)\big(H_{sum}(q_2, p_3) - H_b(p_3)\big)\big]
\end{aligned}
\tag{41}
$$

where $p_3 = (1-p)p_1 + pp_2$ and $\tilde{\eta} = p + \eta_1(1-p)$.

(b) For $\frac{6}{5} \leq \alpha \leq \frac{4}{3}$, we can achieve for any $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$ and $(\eta_1, \gamma_1) \in \left[\frac{1}{2}, 1\right]^2$

$$
\begin{aligned}
R_1 =& \left(2n_d - \tfrac{3n_c}{2}\right) + (1-p)\big[\left(2n_d - \tfrac{3n_c}{2}\right)H_b(\eta_1) + 2(n_c - n_d) + (3n_d - 2n_c)H_b(p_1)\big] \\
&+ p\big[(n_c - n_d)(1 - H_b(q_3)) + (2n_d - \tfrac{3n_c}{2})(1 - H_b(\tilde{\gamma})) + (\tfrac{5n_c}{2} - 3n_d)\big]
\end{aligned}
\tag{42}
$$

where $q_3 = (1-p)q_1 + pq_2$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$
\begin{aligned}
R_2 =& \left(2n_d - \tfrac{3n_c}{2}\right) + (1-p)\big[\left(2n_d - \tfrac{3n_c}{2}\right)H_b(\gamma_1) + 2(n_c - n_d) + (3n_d - 2n_c)H_b(q_1)\big] \\
&+ p\big[(n_c - n_d)(1 - H_b(p_3)) + (2n_d - \tfrac{3n_c}{2})(1 - H_b(\tilde{\eta})) + (\tfrac{5n_c}{2} - 3n_d)\big]
\end{aligned}
\tag{43}
$$

where $p_3 = (1-p)p_1 + pp_2$, and $\tilde{\eta} = p + \eta_1(1-p)$. Remark 7 also applies to the parameters η_1 and γ_1 in (42) and (43).

(c) For $\frac{4}{3} \leq \alpha \leq \frac{3}{2}$, we can achieve for any $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$ and $(\eta_1, \gamma_1) \in \left[\frac{1}{2}, 1\right]^2$,

$$
\begin{aligned}
R_1 =& (n_d - \tfrac{n_c}{2}) + (1-p)\big[(3n_d - 2n_c)(1 + H_b(p_1)) + \left(\tfrac{3n_c}{2} - 2n_d\right)(1 + H_b(\eta_1))\big] \\
&+ p\big[(3n_d - 2n_c)(1 - H_b(q_3)) + (\tfrac{3n_c}{2} - 2n_d)(1 - H_b(\tilde{\gamma}))\big]
\end{aligned}
\tag{44}
$$

$$
\begin{aligned}
R_2 =& (n_d - \tfrac{n_c}{2}) + (1-p)\big[(3n_d - 2n_c)(1 + H_b(q_1)) + \left(\tfrac{3n_c}{2} - 2n_d\right)(1 + H_b(\gamma_1))\big] \\
&+ p\big[(3n_d - 2n_c)(1 - H_b(p_3)) + (\tfrac{3n_c}{2} - 2n_d)(1 - H_b(\tilde{\eta}))\big]
\end{aligned}
\tag{45}
$$

where $q_3 = (1-p)q_1 + pq_2$, $\tilde{\gamma} = p + \gamma_1(1-p)$, $p_3 = (1-p)p_1 + pp_2$ and $\tilde{\eta} = p + \eta_1(1-p)$. Remark 7 also applies to the parameters η_1 and γ_1 in (44) and (45).

(d) For $\frac{3}{2} \leq \alpha \leq 2$, we can achieve for any $\eta_1, \gamma_1 \in \left[\frac{1}{2}, 1\right]$

$$
R_1 = (n_c - n_d) + (1-p)\big[(n_d - \tfrac{n_c}{2})(1 + H_b(\eta_1))\big] + p(n_d - \tfrac{n_c}{2})(1 - H_b(\tilde{\gamma}))
\tag{46}
$$

$$R_2 = (n_c - n_d) + (1 - p)\left[(n_d - \tfrac{n_c}{2})(1 + H_b(\gamma_1))\right] + p(n_d - \tfrac{n_c}{2})(1 - H_b(\tilde{\eta})) \qquad (47)$$

where $\tilde{\gamma} = p + \gamma_1(1 - p)$ and $\tilde{\eta} = p + \eta_1(1 - p)$. Remark 7 also applies to the parameters η_1 and γ_1 in (46) and (47).

In each region, we optimize numerically over the set of parameters, exploiting in some cases that there is symmetry (except for $\alpha = 1$) between the corresponding parameters of both users.

4.3. Local CSIRT vs. Local CSIR

To evaluate the effect of exploiting local CSI at the transmitter side, we plot in Figures 2–4 the converse and achievability bounds for local CSIR and local CSIRT. For each interference region, we choose one value of α. We omit the VWI region because in this region both local CSIR and local CISRT coincide. We observe that for all interference regions, except in the VWI region, local CSIRT outperforms local CSIR. We further observe that the largest improvement is obtained for $p = \frac{1}{2}$. This is not surprising, since in this case the uncertainty about the interference states is the largest.
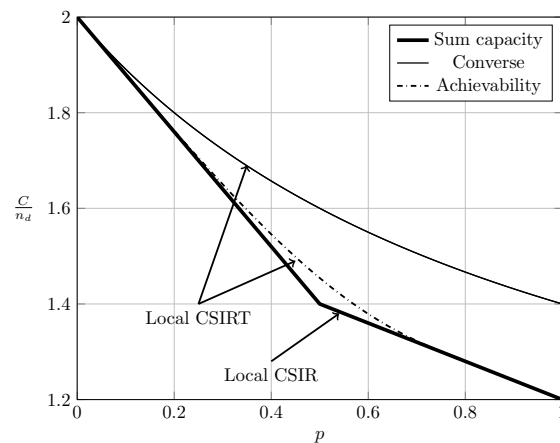

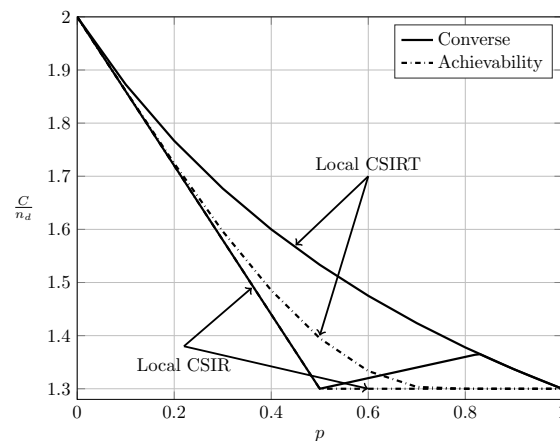**Figure 2.** Local CSIRT vs. local CSIR for $\alpha = \frac{3}{5}$ (WI).

Figure 3. Local CSIRT vs. local CSIR for $\alpha = \frac{7}{10}$ (MI).
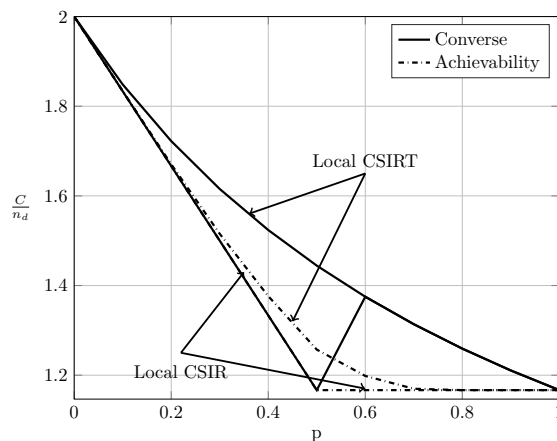
Figure 4. Local CSIRT vs. local CSIR for $\alpha = \frac{7}{6}$ (SI).

4.4. Quasi-Static vs. Ergodic Setup

As observed in the previous subsection, for the ergodic setup local CSIRT outperforms local CSIR in all interference regions (except VWI). In contrast, the opportunistic rates achievable in the quasi-static setup for local CSIRT coincide with those achievable for local CSIR. In other words, the availability of local CSI at the transmitter is only beneficial in the ergodic setup but not in the quasi-static one. This remains to be true even if we consider the average sum capacity rather than the sum rate region. Intuitively, in the coherent setup, the achievable rates depend on the input distributions of X_1^K and X_2^K, and adapting these distributions to the interference state yields a rate gain. In contrast, in the quasi-static setup, we treat the two interference states separately: the worst-case rates are designed for the worst case (where both receivers experience interference), and the opportunistic rates are designed for the best case (where the corresponding receiver is interference-free).

Given that the opportunistic rate region $(R, \Delta R(V_1, V_2))$ is not enhanced by the availability of local CSI at the transmitter, it follows directly that the same is true for the average sum capacity, defined in (23). Note, however, that it is unclear whether (23) corresponds to the best strategy to transmit several messages over independent uses of a quasi-static channel when the transmitters have access to local CSI. Indeed, in this case transmitter i may choose the values for R_i and $\Delta R_i(0)$ as a function of the interference state B_i, potentially giving rise to a larger average sum capacity. Yet, the set of achievable rate pairs $(R_i, \Delta R_i(0))$ depends on the choice of $(R_j, \Delta R_j(0))$ of transmitter $j \neq i$, which transmitter i may not deduce since it has no access to the other transmitter's CSI. How the transmitters should adapt their rates to the interference state remains therefore an open question.

5. Global CSIRT

We next present converse and achievability bounds for global CSIRT. In this scenario, the transmitters may agree on a specific coding scheme that depends on the realization of (B_1^K, B_2^K). This allows for a more elaborated cooperation between the transmitters and strictly increases the sum capacity compared to the local CSIR/CSIRT scenarios.

5.1. Quasi-Static Channel

In the quasi-static scenario with global CSIRT, the messages are, strictly speaking, not opportunistic. Instead, transmitters can choose the message depending on the true state of the interference links, so the strategy is perhaps better described as *rate adaptation*. Nevertheless, the definitions of worst-case sum rate and opportunistic sum rate in Section 2.1 still apply in this case. To keep notation consistent, we use the definition of "opportunism" also for global CSIRT.

5.1.1. Independent Case

Assume first that the sequences B_1 and B_2 are independent of each other.

Theorem 6 (Opportunistic sum capacity for global CSIRT). *Assume that B_1 and B_2 are independent of each other. For $0 < p < 1$, the opportunistic sum capacity region is the union of the set of rate tuples $(R, \Delta R(00), \Delta R(01), \Delta R(10))$ satisfying* (12)–(14) *and*

$$R + \Delta R(00) \le 2n_d \tag{48}$$

$$R + \Delta R(01) \le (n_d - n_c)^+ + \max(n_d, n_c) \tag{49}$$

$$R + \Delta R(10) \le (n_d - n_c)^+ + \max(n_d, n_c). \tag{50}$$

Proof. The converse bounds are proved in Appendix A.5. The achievability bounds are achieved by the following achievability scheme: For $\mathbf{B} = [0,0]$ we use all the n_d sub-channels of both parallel channels. For $\mathbf{B} = [0,1]$ and $\mathbf{B} = [1,0]$ and the VWI/WI regions, we use all n_d sub-channels and the receivers decode them only if they are not affected by interference. For the MI/SI regions, we treat the bursty IC as a non-bursty IC and use the achievability schemes of the IC proposed in [20]. The details can be found in Appendix A.6. □

Remark 8. *The proofs of Theorems 3 and 6 merely require that the joint distribution $p_{b_1 b_2} \triangleq \Pr\{\mathbf{B} = [b_1, b_2]\}$ satisfies $p_{00} < 1$, $p_{01} > 0$, $p_{10} > 0$ and $p_{11} > 0$. Thus, these theorems also apply to the case where B_1 and B_2 are dependent, as long as they are not fully correlated.*

Table 4 summarizes the results of Theorem 6. Observe that for VWI and WI opportunistic messages can be transmitted reliably at a positive rate, while for MI and SI this is only the case if both links are interference-free.

Table 4. Opportunistic sum capacity for global CSIRT when the worst-case sum rate is maximized and B_1 and B_2 are independent.

Rates	VWI	WI	MI	SI
C	$2(n_d - n_c)$	$2n_c$	$2n_d - n_c$	n_c
$\Delta C(00)$	$2n_c$	$2(n_d - n_c)$	n_c	$2n_d - n_c$
$\Delta C(01)/\Delta C(10)$	n_c	$2n_d - 3n_c$	0	0

5.1.2. Fully Correlated Case

Next, we consider the case in which the interference states are fully correlated. In this scenario, local CSIRT coincides with global CSIRT.

Theorem 7 (Opportunistic sum capacity for global CSIRT). *Assume that B_1 and B_2 are fully correlated. For $0 \le p < 1$, the opportunistic sum capacity region is the union of the set of rate pairs $(R, \Delta R(00))$ satisfying* (12)–(14) *and*

$$R + \Delta R(00) \le 2n_d. \tag{51}$$

Proof. For the converse bound, we note that the analysis in Appendix A.5 applies directly to the case where the states B_1 and B_2 are fully correlated, with the only difference that there are only two possible cases $\mathbf{B} = [0,0]$ and $\mathbf{B} = [1,1]$. The result follows then from (A59), (A60) and (A62). For the achievability bound, we use an achievability scheme where the opportunistic messages are only decoded in absence of interference at the intended receiver. In this case, we have two parallel

interference-free channels, for which the optimal strategy consists of transmitting uncoded bits in the n_d sub-channels. □

Table 5 summarizes the results of Theorem 7. Observe that the worst-case sum capacity C and the opportunistic sum capacity $\Delta C(00)$ when the channel is interference-free do not depend on the correlation between B_1 and B_2. The only difference between the independent and fully correlated case is that the interference states $[0,1]$ and $[1,0]$ are impossible if $B_1 = B_2$.

Table 5. Opportunistic sum capacity for global CSIRT when the worst-case sum rate is maximized and B_1 and B_2 are fully correlated.

Rates	VWI	WI	MI	SI
C	$2(n_d - n_c)$	$2n_c$	$2n_d - n_c$	n_c
$\Delta C(00)$	$2n_c$	$2(n_d - n_c)$	n_c	$2n_d - n_c$

5.2. Ergodic Channel

5.2.1. Independent Case

When the sequences B_1^K and B_2^K are independent of each other, we have the following theorems.

Theorem 8 (Converse bounds for global CSIRT). *Assume that B_1^K and B_2^K are independent of each other. The sum rate R for the bursty IC is upper-bounded by*

$$R \leq 2(1-p)n_d + p\left[(n_d - n_c)^+ + \max(n_d, n_c)\right] \tag{52}$$

and

$$R \leq 2\left[p(1-p)\{(n_d - n_c)^+ + \max(n_d, n_c)\} + (1-p)^2 n_d + p^2 \max\{(n_d - n_c)^+, n_c\}\right]. \tag{53}$$

Proof. The proof of (52) follows along similar lines as (18) but noting that, for global CSIRT, X_i^K depends on both B_1^K and B_2^K. The proof of (53) is based on pairing the interference states according the four possible combinations of $(B_{1,k}, B_{2,k})$. See Appendix B.3 for details. □

Remark 9. *The proof of Theorem 8 can be extended to consider an arbitrary joint distribution $p_{b_1 b_2} \triangleq \Pr\{\mathbf{B}_k = [b_1, b_2]\}$. In this case (52) is replaced by*

$$R \leq 2(p_{00} + p_{01})n_d + (p_{10} + p_{11})\left[(n_d - n_c)^+ + \max(n_d, n_c)\right]$$
$$R \leq 2(p_{00} + p_{10})n_d + (p_{01} + p_{11})\left[(n_d - n_c)^+ + \max(n_d, n_c)\right]$$

and (53) becomes

$$R \leq (p_{01} + p_{10})[(n_d - n_c)^+ + \max(n_d, n_c)] + 2\left[p_{00}n_d + p_{11}\max\{(n_d - n_c)^+, n_c\}\right].$$

Theorem 9 (Achievability bounds for global CSIRT). *Assume that B_1^K and B_2^K are independent of each other. The following sum rates R are achievable over the bursty IC:*

$$R = 2\left[p(1-p)(2n_d - n_c) + (1-p)^2 n_d + p^2 \max\{(n_d - n_c)^+, n_c\}\right], \quad \textit{(VWI, WI)} \tag{54}$$

$$R = 4n_d p_{\min} + 2n_d(1-p)^2 + (2n_d - n_c)(2p - p^2 - 3p_{\min}), \quad \textit{(MI)} \tag{55}$$

$$R = 2(n_d + n_c)p_{\min} + 2n_d(1-p)^2 + n_c(2p - p^2 - 3p_{\min}), \quad \textit{(SI)} \tag{56}$$

where $p_{\min} \triangleq \min(p^2, p(1-p))$.

Proof. The sum rate (54) is achieved by using the optimal scheme for the non-bursty IC when any of the two receivers is affected by interference [20], and by using uncoded transmission when there is no interference. The sum rates (55) and (56) are novel. See Appendix B.4 for details.　□

Remark 10. *In contrast to the local CSIR scenario, the achievability schemes presented in Theorem 9 differ noticeably from those in [12] for the binary IC. Indeed, while both works exploit global CSIRT to enable cooperation between users, [12] assumes that only delayed CSI is present. The achievability schemes presented in Theorem 9 thus cannot be applied directly to the model considered in [12].*

Table 6 summarizes the results of Theorems 8 and 9. We write the sum capacity in bold face when converse and achievability bounds coincide. In Table 6, we define

$$\mathbf{C}_{\text{GMI}} \triangleq \min \left\{ 2n_d - pn_c, 2\left[(1 - p^2) - (1 - 2p)\alpha p\right] \right\} \tag{57}$$

$$\mathbf{C}_{\text{GSI}} \triangleq \min \left[n_c p + 2(1 - p)n_d, 2n_d(1 - p)^2 + 2n_c p \right] \tag{58}$$

where "G" stands for "global CSIRT".

Table 6. Bounds on the sum capacity C for global CSIRT when B_1^K and B_2^K are independent.

Regions	Achievability	Converse
VWI	$\mathbf{2(n_d - pn_c)}$	
WI	$\mathbf{2[(1 - p^2)n_d + (1 - 2p)pn_c]}$	
MI	$4n_d p_{\min} + 2n_d(1 - p)^2 + (2n_d - n_c)(2p - p^2 - 3p_{\min})$	$\mathbf{C}_{\text{GMI}}$
SI	$2(n_d + n_c)p_{\min} + 2n_d(1 - p)^2 + n_c(2p - p^2 - 3p_{\min})$	$\mathbf{C}_{\text{GSI}}$

5.2.2. Fully Correlated Case

We next discuss the case where the sequences B_1^K and B_2^K are fully correlated, i.e., $B_1^K = B_2^K$.

Theorem 10 (Converse bounds for global CSIRT). *Assume that B_1^K and B_2^K are fully correlated. The sum rate R for the bursty IC is upper-bounded by*

$$R \leq 2(1 - p)n_d + p\{(n_d - n_c)^+ + \max(n_d, n_c)\} \tag{59}$$

$$R \leq 2\left[(1 - p)n_d + p\max\{(n_d - n_c)^+, n_c\}\right]. \tag{60}$$

Proof. The proof of (59) follows similar steps as in Appendix B.3.1 but considering $B_1^K = B_2^K = B^K$. The proof of (60) is given in Appendix B.5. See also Remark 9.　□

Theorem 11 (Achievability bounds for global CSIRT). *Assume that B_1^K and B_2^K are fully correlated. The following sum rates R are achievable over the bursty IC:*

$$R = 2\left[(1 - p)n_d + p\max\{(n_d - n_c)^+, n_c\}\right], \quad \textit{VWI/WI} \tag{61}$$

$$R = 2(1 - p)n_d + p\{(n_d - n_c)^+ + \max(n_d, n_c)\}, \quad \textit{MI/SI}. \tag{62}$$

Proof. The sum rates (61) and (62) are achieved by using the optimal scheme for the non-bursty IC when the two receivers are affected by interference [20], and by using uncoded transmission in absence of interference.　□

Table 7 summarizes the results of Theorems 10 and 11. For global CSIRT and fully correlated B_1^K and B_2^K, converse and achievability bounds coincide. Thus, (61) and (62) indicate the sum capacity.

Table 7. Bounds on the sum capacity C for global CSIRT when B_1^K and B_2^K are fully correlated.

Regions	Bounds
VWI	$2(n_d - pn_c)$
WI	$2[(1-p)n_d + pn_c]$
MI	$2(1-p)n_d + p(2n_d + n_c)$
SI	$2(1-p)n_d + p(n_c)$

5.3. Quasi-Static vs. Ergodic Setup

Similar to the average sum capacity for local CSIR defined in Section 3.3, we define the average sum capacity for global CSIRT when B_1 and B_2 are independent as

$$\bar{C} = p^2 \sup_R \{R\} + p(1-p) \sup_{(R,\Delta R(01))} \{R + \Delta R(01)\} + p(1-p) \sup_{(R,\Delta R(10))} \{R + \Delta R(10)\}$$
$$+ (1-p)^2 \sup_{(R,\Delta R(00))} \{R + \Delta R(00)\} \tag{63}$$

where the suprema are over all rate tuples $(R, \Delta R(00), \Delta R(01), \Delta R(10))$ that satisfy Theorems 2 and 6. The intuition behind (63) is the same as that behind (23) for local CSIR, but with global CSIRT the transmitters can adapt their rates $(R_i, \Delta R_i(V_i))$ to the interference state. For example, the first term on the right-hand side (RHS) of (63) corresponds to the interference state $[1, 1]$, in which case we transmit at total sum rate R; the second term corresponds to the interference state $[0, 1]$, in which case we transmit at total sum rate $R + \Delta R(01)$; and so on.

Table 8 summarizes the average sum capacity for the different interference regions. The average sum capacities for VWI and WI coincide with the sum capacities in the ergodic setup (see Table 6). In contrast, for MI and SI, the average sum capacities are smaller than the sum capacities in the ergodic setup.

Table 8. Average sum capacity when B_1 and B_2 are independent.

Regions	Bounds
VWI	$2(n_d - pn_c)$
WI	$2[(1-p^2)n_d + (1-2p)pn_c]$
MI	$2n_d - pn_c(2-p)$
SI	$2n_d(1-p)^2 + pn_c(2-p)$

Similarly, in the fully correlated case, we define the average sum capacity as

$$\bar{C} \triangleq p \sup_R \{R\} + (1-p) \sup_{(R,\Delta R(00))} \{(R + \Delta R(00))\} \tag{64}$$

where the suprema are over all rate pairs $(R, \Delta R(00))$ that satisfy Theorems 2 and 7. The corresponding results are summarized in Table 9.

Table 9. Average sum capacity when B_1 and B_2 are fully correlated.

Regions	Bounds
VWI	$2(n_d - pn_c)$
WI	$2[(1-p)n_d + pn_c]$
MI	$2(1-p)n_d + p(2n_d + n_c)$
SI	$2(1-p)n_d + p(n_c)$

We observe that the average sum capacities coincide with the sum capacities of the ergodic setup.

6. Exploiting CSI

In this section, we study how the level of CSI affects the sum rate in the quasi-static and ergodic setups.

For the quasi-static channel, Figures 5 and 6 show the total sum capacity presented in Theorems 3, 6 and 7. Specifically, we plot the normalized total sum capacity $\frac{C+\Delta C}{n_d}$ versus α, comparing scenarios of local CSIR/CSIRT and global CSIRT. We analyze separately the cases $\mathbf{B} = [0,0]$ and $\mathbf{B} = [0,1]$. For the case where $\mathbf{B} = [0,0]$ and global CSIRT, the total sum capacity is $2n_d$ for all interference regions. For $\mathbf{B} = [0,0]$ and local CSIR/CSIRT, the total sum capacity is $2n_d$ for VWI and VSI, but is strictly smaller in the remaining interference regions. Hence, in these regions global CSIRT outperforms local CSIR/CSIRT. For the case where $\mathbf{B} = [0,1]$, the total sum capacity is equal to $(n_d - n_c)^+ + \max(n_d, n_c)$ irrespective of the level of CSI.
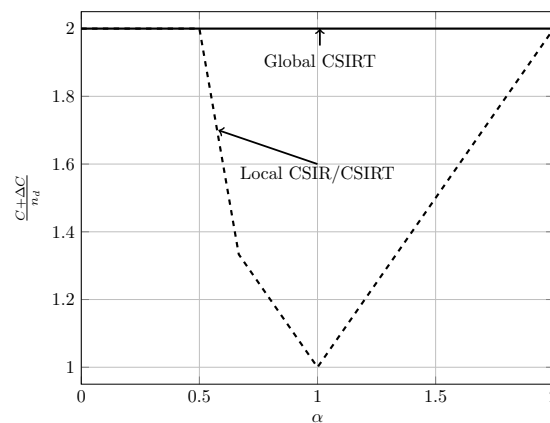

Figure 5. Total sum capacity for $\mathbf{B} = [0,0]$, for local CSIR/CSIRT and global CSIRT.
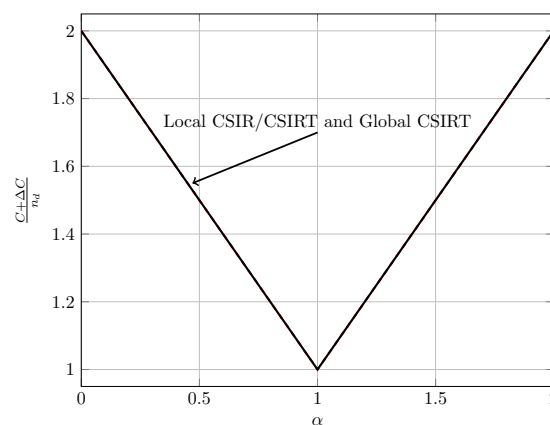

Figure 6. Total sum capacity for $\mathbf{B} = [0,1]$, for local CSIR/CSIRT and global CSIRT.

We further observe that the opportunistic-capacity region for local CSIRT is equal to that for local CSIR. Thus, local CSI at the transmitter is not beneficial. As we shall see later, this is in stark contrast to the ergodic setup, where local CSI at the transmitter-side is beneficial. Intuitively, in the ergodic case the input distributions of X_1^K and X_2^K depend on the realizations of B_1^K and B_2^K, respectively. Hence, adapting the input distributions to these realizations increases the sum capacity. In contrast, in the quasi-static case, the worst-case scenario (presence of interference) and the best-case scenario (absence of interference) are treated separately. Hence, there is no difference to the case of local CSIR.

For the ergodic setup, Figures 7–10 show the converse and achievability bounds presented in Theorems 4, 5, 8 and 9. We further include the results on local CSIRT presented in Section 4. Specifically, we plot the normalized sum capacity $\frac{C}{n_d}$ versus the probability of presence of interference p, comparing scenarios of local CSIR, local CSIRT and global CSIRT when B_1^K and B_2^K are independent of each other. The shadowed areas correspond to the regions where achievability and converse bounds do not coincide.
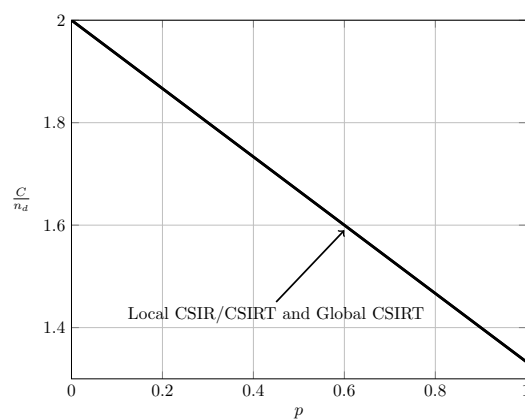

Figure 7. Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{1}{3}$ (VWI).
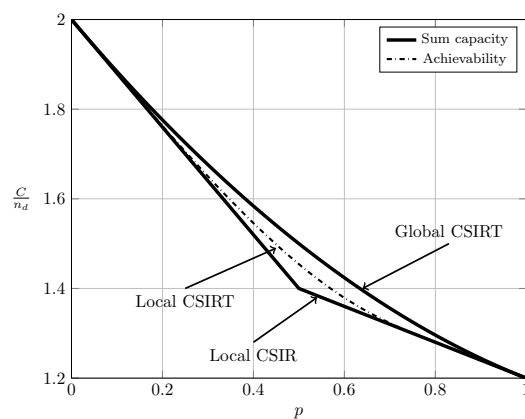

Figure 8. Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{3}{5}$ (WI).
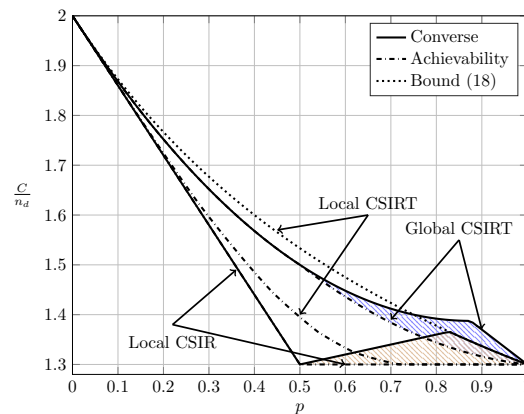
Figure 9. Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{7}{10}$ (MI).
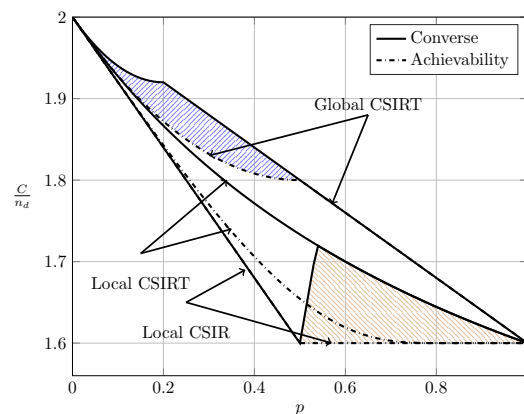

Figure 10. Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{8}{5}$ (SI).

Figure 7 reveals that in the VWI region the sum capacity is equal to $2(n_d - pn_c)$, irrespective of the availability of CSI (see Figure 7). Thus, in this region access to global CSIRT is not beneficial compared to the local CSIR scenario. In the VSI region, the sum capacity of the non-bursty IC is equal to $2n_d$, which is that of two parallel channels without interference [15] (Section II-A). Therefore, burstiness of the interference (and hence CSI) does not affect the sum capacity.

In the WI region, shown in Figure 8, the converse and achievability bounds for local CSIR and global CSIRT coincide and it is apparent that global CSIRT outperforms local CSIR. In the MI and SI regions, the converse and achievability bounds only coincide for certain regions of p. Nevertheless, Figures 9 and 10 show that, in almost all cases, global CSIRT outperforms local CSIR. (For the case presented in Figure 9 $\left(\alpha = \frac{7}{10}\right)$, we also present the local CSIRT converse bound (18), although it is looser for some values of p, with respect to the one depicted for global CSIRT.) Local CSIRT outperforms local CSIR in all interference regions (except VWI). We stress again the fact that this was not the case in the quasi-static scenario, where both coincide.

We next consider the case where B_1^K and B_2^K are fully correlated. For this scenario, [7,23] studied the effect of perfect feedback on the bursty IC. For comparison, the non-bursty IC with feedback was studied by Suh et al. in [25], where it was demonstrated that the gain of feedback becomes arbitrarily large for certain interference regions (VWI and WI) when the signal-to-noise-ratio increases. This gain corresponds to a better resource utilization and thereby a better resource sharing between users. Specifically, [7,23] (bursty IC) and [25] (non-bursty IC) assume that noiseless, delayed feedback

is available from receiver i to transmitter i ($i = 1, 2$). For the symmetric setup treated in this paper, [7] (Theorem 3.2) or [23] (Theorem 3.2) showed the following:

Theorem 12 (Channel capacity for the bursty IC with feedback [7,23])**.** *The sum capacity of the bursty IC with noiseless, delayed feedback is given by*

$$C = \begin{cases} 2n_d - 2\frac{p}{1+p}n_c, & \alpha \leq 1, \\ 2\frac{1-p}{1+p}n_d + 2\frac{p}{1+p}n_c & 1 < \alpha \leq 2, \\ 2(1-p)n_d + pn_c, & 2 < \alpha. \end{cases} \tag{65}$$

Proof. See [7] (Sections IV and V), [23] (Sections IV and V, Appendices A, C, D). $\square$

Observe that (65) for $\alpha \leq 2$ coincides with (18). This implies that local CSIRT can never outperform delayed feedback. Intuitively, feedback contains not only information about the channel state, but also about the previous symbols transmitted by the other transmitter, which can be exploited to establish a certain cooperation between the transmitters. Figures 11–14 show the bounds on the normalized sum capacity, $\frac{C}{n_d}$, comparing the scenarios of local CSIR versus global CSIRT when the interference states are fully correlated, i.e., $B_1^K = B_2^K$. They further show the sum capacity for the case where the transmitters have noiseless delayed feedback [7]. The shadowed areas correspond to the regions where achievability and converse bounds do not coincide.
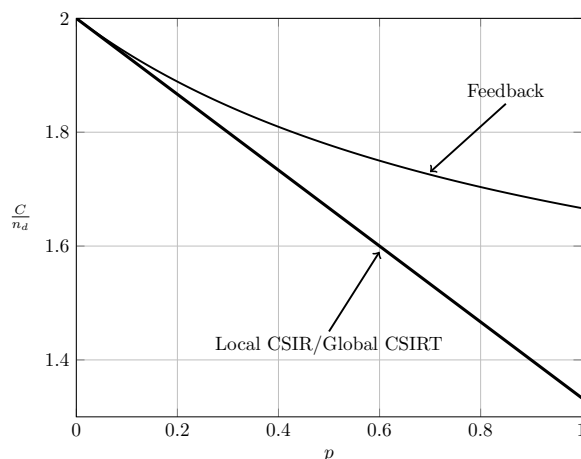

Figure 11. Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{1}{3}$ (VWI).
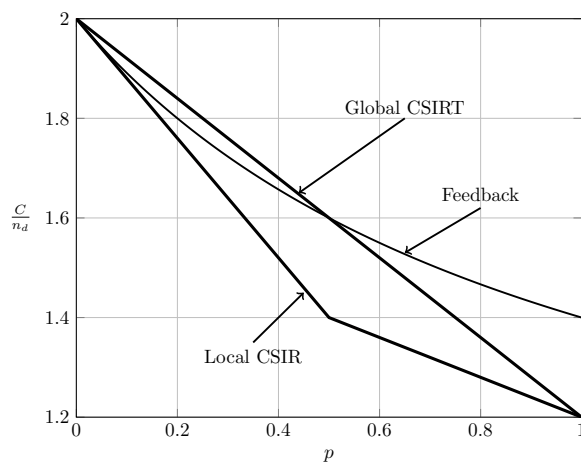

Figure 12. Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{3}{5}$ (WI).
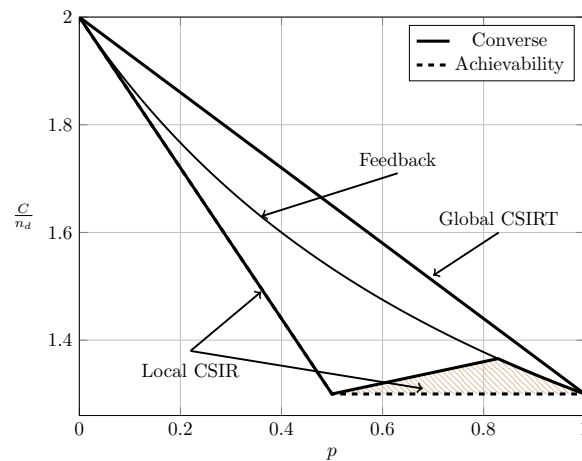
Figure 13. Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{7}{10}$ (MI).
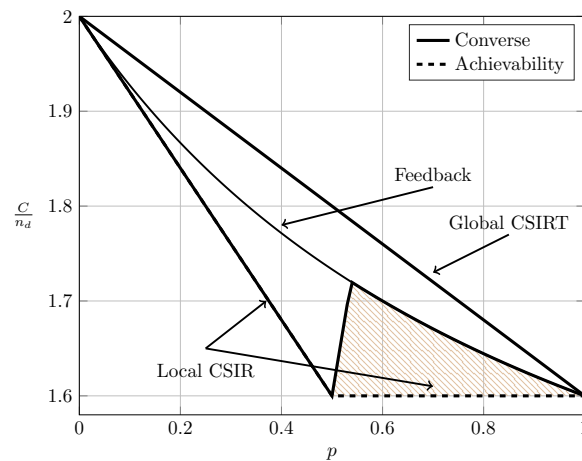

Figure 14. Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{8}{5}$ (SI).

Figure 11 reveals that feedback in the VWI region outperforms the non-feedback case, irrespective of the availability of CSI. Wang et al. [7] have further shown that feedback also outperforms the non-feedback case in the VSI region. The order between global CSIRT and the feedback scheme is not obvious. There are regions where global CSIRT outperforms the feedback scheme and vice versa. Indeed, on the one hand, feedback contains information about the previous interference states and previous symbols transmitted by the other transmitter, permitting the resolution of collisions in previous transmissions. On the other hand, global CSIRT provides *non-causal* information about the interference states, allowing a better adaptation of the transmission strategy to the interference burstiness.

7. Exploiting Interference Burstiness

To better illustrate the benefits of interference burstiness, we show the normalized sum capacity as a function of α, in order to appreciate all the interference regions. In the non-bursty IC ($p = 1$), this curve corresponds to the well-known W-curve obtained by Etkin et al. in [26]. We next study how burstiness affects this curve in the different considered scenarios.

In the quasi-static setup, burstiness can be exploited by sending opportunistic messages. We consider the total sum capacity for the case where the worst-case rate R is maximized. For local CSIR/CSIRT, Theorem 3 suggests that the use of an opportunistic code is only beneficial if the interference region is VWI or WI. For other interference regions there is no benefit. In contrast, for

global CSIRT an opportunistic code is beneficial for all interference regions (except for VSI where the sum capacity corresponds to that of two parallel channels without interference).

Figures 15 and 16 illustrate these observations. Specifically, in Figures 15 and 16 we show the normalized total sum capacity achieved under local CSIR/CSIRT and global CSIRT when the interference states are independent. We observe that, for local CSIR, the opportunistic rates $\Delta R_1(0)$ and $\Delta R_2(0)$, are only positive in the VWI and WI regions. In these regions, if only one of the receivers is affected by interference the sum capacity is given by the worst-case rate R plus one opportunistic rate of the user which is not affected by interference. In absence of interference at both receivers, both receivers can decode opportunistic messages. Hence, the total sum capacity is equal to $C + \Delta C_1(0) + \Delta C_2(0)$. For global CSIRT we can observe that, when only one of the receivers is affected by interference, we achieve the same total sum capacity as in the local CSIR/CSIRT. However, in absence of interference at both receivers, we achieve the trivial upper bound corresponding to two parallel channels. The fully correlated scenario can be considered as a subset of the independent scenario. Indeed, for the case $\mathbf{B} = [0,0]$ and $\mathbf{B} = [1,1]$ we obtain the same total sum capacity as for the independent scenario. The main difference is that in the fully correlated scenario the interference states $\mathbf{B} = [0,1]$ and $\mathbf{B} = [1,0]$ are impossible.

Figure 15. Normalized total sum capacity $\frac{C+\Delta C}{n_d}$ as a function of α for local CSIR/CSIRT when B_1 and B_2 are independent.

Figure 16. Normalized total sum capacity $\frac{C+\Delta C}{n_d}$ as a function of α for global CSIRT when B_1 and B_2 are independent.

For the ergodic case, Figures 17 and 18 show the bounds on the normalized sum capacity, $\frac{C}{n_d}$, as a function of α when B_1^K and B_2^K are independent. The shadowed areas correspond to the regions where

achievability and converse bounds do not coincide. We further show the W-curve. Observe that for $p \leq \frac{1}{2}$ the sum capacity as a function of α forms a V-curve instead of the W-curve. Further observe how the sum capacity approaches the W-curve as p tends to one.
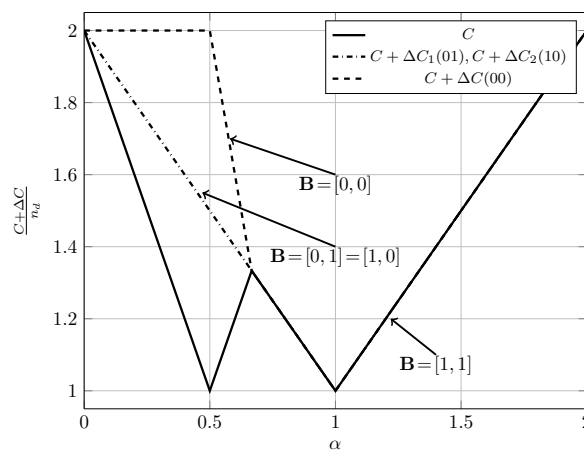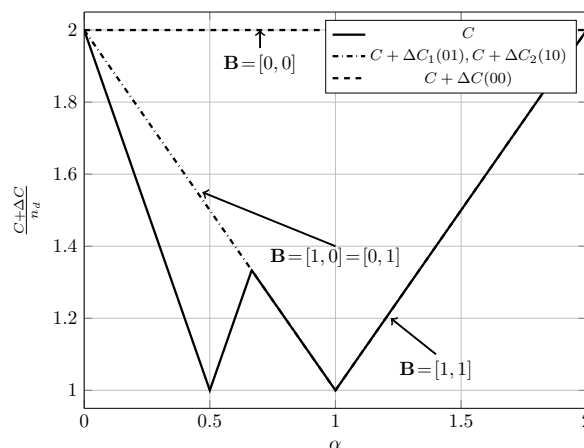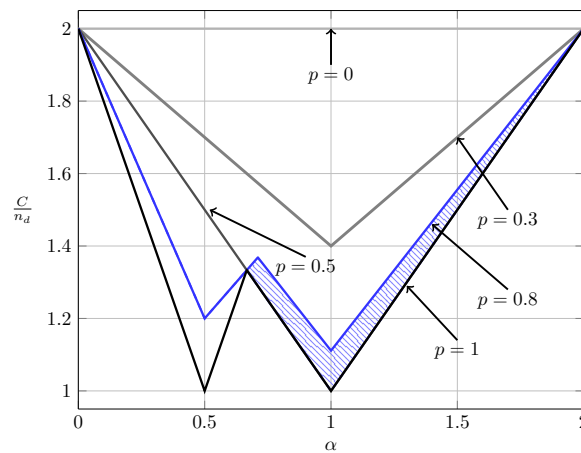

Figure 17. Normalized sum capacity $\frac{C}{n_d}$ as a function of α for local CSIR/CSIRT when B_1^K and B_2^K are independent.
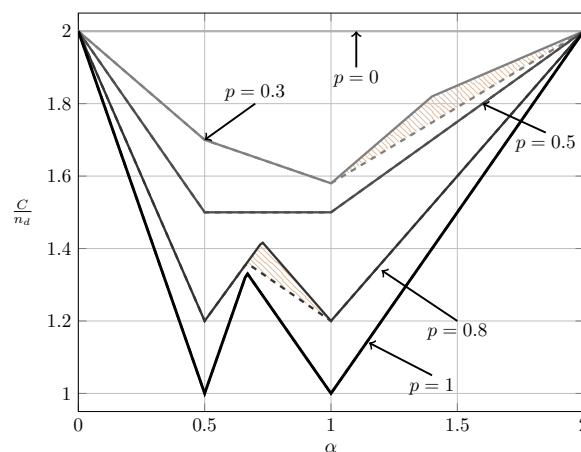
Figure 18. Normalized sum capacity $\frac{C}{n_d}$ as a function of α for global CSIRT when B_1^K and B_2^K are independent.

In Figure 19 we show the bounds on the normalized sum capacity, $\frac{C}{n_d}$, as a function of α for global CSIRT when B_1^K and B_2^K are fully correlated. (For local CSIR the sum capacity is not affected by the correlation between B_1^K and B_2^K, so the curve for $\frac{R}{n_d}$ as a function of α coincides with the one obtained in Figure 17.) We observe that, for all values of $p > 0$, the sum capacity forms a W-curve similar to the W-curve for $p = 1$. This is the case because, when both interference states are fully correlated, the bursty IC is a combination of an IC and two parallel channels.

We observe that for global CSIRT the burstiness of the interference is beneficial for all interference regions and all values of p. For local CSIR, burstiness is beneficial for all values of p for VWI and WI. However, for MI and SI, burstiness is only of clear benefit for $p \leq \frac{1}{2}$. It is yet unclear whether burstiness is also beneficial in these interference regions when $p > \frac{1}{2}$. To shed some light on this question, note that evaluating the converse bound in [23] (Lemma A.1), which yields (21), for inputs X_1^K and X_2^K that are temporally independent, we recover the achievability bound (20). Since for MI/SI and $p \geq \frac{1}{2}$ this bound coincides with the rates achievable over the non-bursty IC, this implies that an achievability scheme can only exploit the burstiness of the interference in this regime if it introduces

some temporal correlation (this observation is also revealed by considering the average sum capacity for the quasi-static case). In fact, for global CSIRT the achievability schemes proposed in Theorem 9 for MI and SI copy the same bits over several coherence blocks, i.e., they exhibit a temporal correlation, which cannot be achieved using temporally independent distributions. However, the temporal pattern of these bits requires knowledge of both interference states, so this approach cannot be adapted to the cases of local CSIR/CSIRT. In contrast, for global CSIRT in the fully correlated case where converse and achievability bounds coincide, it is not necessary to introduce temporal memory. This scenario is simpler, since in this case the channel exhibits only two channel states, a non-bursty IC and two parallel channels.

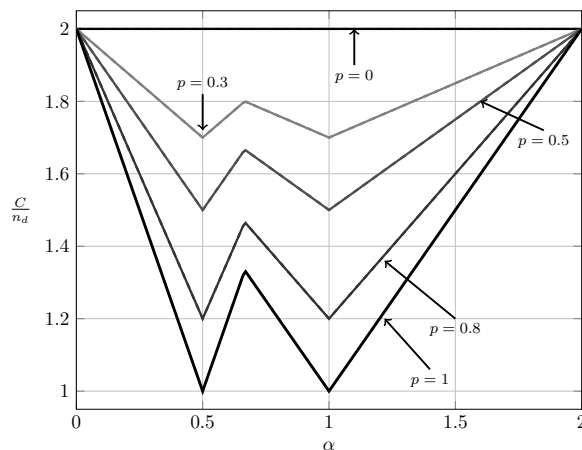**Figure 19.** Normalized sum capacity $\frac{C}{n_d}$ as a function of α for global CSIRT when $B_1^K = B_2^K$.

8. Summary and Conclusions

In this work, we considered a two-user bursty IC in which the presence/absence of interference is modeled by a block-i.i.d. Bernoulli process while the power of the direct and cross links remains constant during the whole transmission. This scenario corresponds, e.g., to a slow-fading scenario in which all the nodes can track the channel gains of the different links, but where the interfering links are affected by intermittent occlusions due to some physical process. While this model may appear over-simplified, it yields a unified treatment of several aspects previously studied in the literature and gives rise to several new results on the effect of the CSI in the achievable rates over the bursty IC. Our channel model encompasses both the quasi-static scenario studied in [3,5] and the ergodic scenario (see, e.g., [7,12]). While the model recovers several cases studied in the literature, it also presents scenarios which have not been previously analyzed. This is the case, for example, for the ergodic setup with local and global CSIRT. Our analysis in these scenarios does not yield matching upper and lower bounds for all interference and burstiness levels. Yet, examining the obtained results, we observe that the best strategies in these scenarios often require elaborated coding strategies for both users that feature memory across different interference. This fact probably explains why no previous results exist in these scenarios. Furthermore, several of our proposed achievability schemes require complex correlation among signal levels. Thus, while the LDM in general provides insights on the Gaussian IC, the proposed schemes may actually be difficult to convert to the Gaussian case.

In the quasi-static scenario, the highest sum rate R that can be achieved is limited by the worst realization of the channel and thus coincides with that of the (non-bursty) IC. We can however transmit at an increased (opportunistic) sum rate $R + \Delta R$ when there is no interference at any of the interfering links. For the ergodic setup, we showed that an increased rate can be obtained when local CSI is present at both transmitter and receiver, compared to that obtained when CSI is only available at the receiver side. This is in contrast to the quasi-static scenario, where the achievable rates for local CSIR and local CSIRT coincide. Featuring global CSIRT at all nodes yields an increased sum rate for both

the quasi-static and the ergodic scenarios. In the quasi-static channel, global CSI yields increased opportunistic rates in all the regions except in the very strong interference region, which is equivalent to having two parallel channels with no interference.

Both in the quasi-static and ergodic scenarios, global CSI exploits interference burstiness for all interference regions (except for very strong interference), irrespective of the level of burstiness. When local CSI is available only at the receiver side, interference burstiness is of clear benefit if the interference is either weak or very weak, or if the channel is ergodic and interference is present at most half of the time. When local CSI is available at each transmitter and receiver and the channel is ergodic, interference burstiness is beneficial in all interference regions except in the very weak and very strong interference regions.

In order to compare the achievable rates of the quasi-static and ergodic setup, one can define the average sum rate of the quasi-static setup for local CSIR/CSIRT as $R + (1 - p)(\Delta R_1(0) + \Delta R_2(0))$, with a similar definition for the average sum rate for global CSIRT. The average sum rate corresponds to a scenario where several codewords are transmitted over independent quasi-static bursty ICs. This, in turn, could be the case if a codeword spans several coherence blocks, but no coding is performed over these blocks. This is in contrast to the ergodic setup where coding is typically performed over different coherence blocks. By the law of large numbers, roughly a fraction of p codewords experiences interference, the remaining codewords are transmitted free of interference. Consequently, an opportunistic transmission strategy achieves the rate $pR + (1 - p)(R + \Delta R_1(0) + \Delta R_2(0))$, which corresponds to the average sum rate. Our results demonstrate that, for local CSIR, the average sum capacity, obtained by maximizing the average sum rate over all achievable rate pairs $(R, \Delta R_1(0) + \Delta R_2(0))$, coincides with the achievable rates in the ergodic setup for all interference regions. In contrast, for local CSIRT, the average sum capacity is strictly smaller than the sum capacity in the ergodic setup. For global CSIRT, average sum capacity and sum capacity coincide for all interference regions when the interference states are fully correlated, and they coincide for VWI and WI when the interference states are independent. For global CSIRT, MI/SI, and independent interference states, the average sum capacity is smaller than the sum capacity in the ergodic setup. In general, the average sum capacity defined for the quasi-static setup never exceeds the sum capacity in the ergodic setup. This is perhaps not surprising if we recall that the average sum capacity corresponds to the case where no coding is performed over coherence blocks. Interestingly, the average sum capacity is not always achieved by maximizing the worst-case rate. For small values of p, it is beneficial to reduce the worst-case rate in order to achieve a larger opportunistic rate.

In our work we considered both the case where the interference states of the two users are independent and the case where the interference states are fully correlated. In both ergodic and quasi-static setups, the results for local CSIR are independent of the correlation between interference states. For other CSI levels, dependence between the interference states helps in all interference regions except very weak and very strong interference regions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs for the Quasi-Static Case

We define $p_{\mathbf{b}} = \Pr\{\mathbf{B} = \mathbf{b}\}$. Clearly, when B_1, B_2 are independent, we have $p_{00} = (1-p)^2$, $p_{11} = p^2$ and $p_{01} = p_{10} = p(1-p)$, and when B_1, B_2 are fully correlated $p_{00} = 1-p$, $p_{11} = p$ and $p_{01} = p_{10} = 0$.

The converse bounds in the quasi-static case are based on an information density approach [27]. In particular, we define the information densities for the bursty IC

$$i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) \triangleq i_{\mathbf{X}_1^N \mathbf{Y}_1^N | \mathbf{B}}(\mathbf{x}_1^N; \mathbf{y}_1^N | \mathbf{b}) = \log \frac{P_{\mathbf{Y}_1^N | \mathbf{X}_1^N, \mathbf{B}}(\mathbf{y}_1^N | \mathbf{x}_1^N, \mathbf{b})}{P_{\mathbf{Y}_1^N | \mathbf{B}}(\mathbf{y}_1^N | \mathbf{b})} \tag{A1}$$

$$i_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b}) \triangleq i_{\mathbf{X}_2^N \mathbf{Y}_2^N | \mathbf{B}}(\mathbf{x}_2^N; \mathbf{y}_2^N | \mathbf{b}) = \log \frac{P_{\mathbf{Y}_2^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{y}_2^N | \mathbf{x}_2^N, \mathbf{b})}{P_{\mathbf{Y}_2^N | \mathbf{B}}(\mathbf{y}_2^N | \mathbf{b})}. \tag{A2}$$

Here and throughout the appendices, we use the notations $\mathbf{X}_i^N = X_i$, $\mathbf{x}_i^N = x_i$, $\mathbf{Y}_i^N = Y_i$, and $\mathbf{y}_i^N = y_i$ to highlight the fact that, in the quasi-static setting, we transmit N symbols in one coherence block.

We further consider the individual error events

$$\mathcal{E}_i(\Gamma_i) \triangleq \left\{ \frac{1}{N} i_i(\mathbf{x}_i^N, \mathbf{y}_i^N, \mathbf{b}) \leq \Gamma_i \right\}, \ i = 1, 2 \tag{A3}$$

and the joint error event

$$\mathcal{E}_{12}(\Gamma) \triangleq \left\{ \frac{1}{N} \left(i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) + i_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b}) \right) \leq \Gamma \right\}. \tag{A4}$$

The proofs of the converse results are based on the following lemmas.

Lemma A1 (Verdú-Han lemma). *Every* (N, R, P_e) *code over a channel* $P_{\mathbf{Y}^N | \mathbf{X}^N}$ *satisfies*

$$P_e \geq \Pr\left\{ \frac{1}{N} i_{\mathbf{X}^N \mathbf{Y}^N}(\mathbf{X}^N; \mathbf{Y}^N) \leq R - \gamma \right\} - e^{-\gamma N} \tag{A5}$$

for every $\gamma > 0$, *where* $\mathbf{X}^N$ *places probability mass* $\frac{1}{2^{NR}}$ *on each codeword and* $i_{\mathbf{X}^N \mathbf{Y}^N}(\mathbf{X}^N; \mathbf{Y}^N) \triangleq \log \frac{P_{\mathbf{Y}^N | \mathbf{X}^N}(\mathbf{y}^N | \mathbf{x}^N)}{P_{\mathbf{Y}^N}(\mathbf{y}^N)}$.

Proof. See [27] (Theorem 4). □

Lemma A2. *Suppose that* $\Pr\{\mathcal{E}_{12}(\Gamma) | \mathbf{B} = \mathbf{b}\} \to 0$ *as* $N \to \infty$. *Then, for each pair* $\mathbf{b} \in \{0, 1\}^2$, *the threshold* Γ *must satisfy the following conditions:*

- *For* $\mathbf{B} = [0, 0]$, Γ *satisfies*

$$\Gamma \leq 2n_d. \tag{A6}$$

- *For* $\mathbf{B} = [0, 1]$ *and* $\mathbf{B} = [1, 0]$, Γ *satisfies* (A6) *and*

$$\Gamma \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A7}$$

- *For* $\mathbf{B} = [1, 1]$, Γ *satisfies* (A6) *and* (A7), *and*

$$\Gamma \leq 2 \max\{(n_d - n_c)^+, n_c\}. \tag{A8}$$

Proof. See Appendix C. □

Appendix A.1. Proof of Theorem 2

In this section we prove the IC channel converse bounds for $p > 0$. This proof assumes global CSIRT, hence the resulting bounds also apply to local CSIR and local CSIRT. Let $P_e^{(N)} = \Pr\{(\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2)\}$, and let us denote by $P_{e_1}^{(N)}$ and $P_{e_2}^{(N)}$ the error probabilities at decoders one and two, respectively:

$$P_{e1}^{(N)} \triangleq \Pr\{\hat{W}_1 \neq W_1\}, \tag{A9}$$

$$P_{e2}^{(N)} \triangleq \Pr\{\hat{W}_2 \neq W_2\}. \tag{A10}$$

Clearly, the error probabilities $P_e^{(N)}$, $P_{e1}^{(N)}$ and $P_{e2}^{(N)}$ are related by the following sets of inequalities

$$\max\left(P_{e1}^{(N)}, P_{e2}^{(N)}\right) \leq P_e^{(N)} \leq P_{e1}^{(N)} + P_{e2}^{(N)} \leq 2\max\left(P_{e1}^{(N)}, P_{e2}^{(N)}\right). \tag{A11}$$

Using these inequalities we conclude that

$$P_e^{(N)} \geq \frac{1}{2}\left(P_{e1}^{(N)} + P_{e2}^{(N)}\right). \tag{A12}$$

We now rewrite (A9) and (A10) as

$$P_{e1}^{(N)} = \sum_{\mathbf{b}} p_{\mathbf{b}} \Pr\{\hat{W}_1 \neq W_1 | \mathbf{B} = \mathbf{b}\}, \tag{A13}$$

$$P_{e2}^{(N)} = \sum_{\mathbf{b}} p_{\mathbf{b}} \Pr\{\hat{W}_2 \neq W_2 | \mathbf{B} = \mathbf{b}\} \tag{A14}$$

and apply the Verdú-Han lemma (Lemma A1) to each of the probability terms $\Pr\{\hat{W}_i \neq W_i | \mathbf{B} = \mathbf{b}\}$, $i = 1, 2$, in (A13) and (A14). This yields

$$\Pr\{\hat{W}_1 \neq W_1 | \mathbf{B} = \mathbf{b}\} \geq \Pr\left\{\frac{1}{N} i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) \leq R_1 - \gamma_1 | \mathbf{B} = \mathbf{b}\right\} - e^{-\gamma_1 N}, \tag{A15}$$

$$\Pr\{\hat{W}_2 \neq W_2 | \mathbf{B} = \mathbf{b}\} \geq \Pr\left\{\frac{1}{N} i_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b}) \leq R_2 - \gamma_2 | \mathbf{B} = \mathbf{b}\right\} - e^{-\gamma_2 N}. \tag{A16}$$

We set $\Gamma_i = R_i - \gamma_i$ and $\Gamma = \Gamma_1 + \Gamma_2 = R - \gamma_1 - \gamma_2$. Then, using the definition of $\mathcal{E}_i$ in (A3), we can write (A15) and (A16) as

$$\Pr\{\hat{W}_1 \neq W_1 | \mathbf{B} = \mathbf{b}\} \geq \Pr\{\mathcal{E}_1(\Gamma_1) | \mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N}, \tag{A17}$$

$$\Pr\{\hat{W}_2 \neq W_2 | \mathbf{B} = \mathbf{b}\} \geq \Pr\{\mathcal{E}_2(\Gamma_2) | \mathbf{B} = \mathbf{b}\} - e^{-\gamma_2 N}. \tag{A18}$$

Comparing the joint error event $\mathcal{E}_{12}(\Gamma)$ in (A4) with $\mathcal{E}_1(\Gamma_1)$ and $\mathcal{E}_2(\Gamma_2)$ in (A3), it can be shown that

$$\mathcal{E}_1(\Gamma_1) \cap \mathcal{E}_2(\Gamma_2) \subseteq \mathcal{E}_{12}(\Gamma), \tag{A19}$$

$$\mathcal{E}_1^c(\Gamma_1) \cap \mathcal{E}_2^c(\Gamma_2) \subseteq \mathcal{E}_{12}^c(\Gamma) \Rightarrow \mathcal{E}_{12}(\Gamma) \subseteq \mathcal{E}_1(\Gamma_1) \cup \mathcal{E}_2(\Gamma_2). \tag{A20}$$

Using (A20) and the union bound, we thus obtain

$$\Pr\{\mathcal{E}_{12}(\Gamma) | \mathbf{B} = \mathbf{b}\} \leq \Pr\{\mathcal{E}_1(\Gamma_1) \cup \mathcal{E}_2(\Gamma_2) | \mathbf{B} = \mathbf{b}\}$$
$$\leq \Pr\{\mathcal{E}_1(\Gamma_1) | \mathbf{B} = \mathbf{b}\} + \Pr\{\mathcal{E}_2(\Gamma_2) | \mathbf{B} = \mathbf{b}\}. \tag{A21}$$

Combining this result with (A12), (A17) and (A18) gives

$$
\begin{aligned}
P_e^{(N)} &\geq \frac{1}{2}\left(P_{e1}^{(N)} + P_{e2}^{(N)}\right) \\
&\geq \frac{1}{2}\sum_{\mathbf{b}} p_{\mathbf{b}}\left(\Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B}=\mathbf{b}\} + \Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B}=\mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\right) \\
&\geq \frac{1}{2}\sum_{\mathbf{b}} p_{\mathbf{b}}\left(\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B}=\mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\right).
\end{aligned}
\tag{A22}
$$

The remainder of this section is devoted to an analysis of $\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B}=\mathbf{b}\}$. Indeed, by (A22) we have for any $\gamma_1, \gamma_2 > 0$ that

$$
\lim_{N\to\infty} P_e^{(N)} \geq \lim_{N\to\infty} \frac{1}{2}\left[p_{11}\epsilon_{11} + p_{00}\epsilon_{00} + p_{10}\epsilon_{10} + p_{01}\epsilon_{01}\right],
\tag{A23}
$$

where $\epsilon_{\mathbf{b}} \triangleq \Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B}=\mathbf{b}\}$. When $p > 0$, the probability p_{11} is strictly positive both when (B_1, B_2) are independent and when they are fully correlated. Since $p_{\mathbf{b}}$ does not depend on N, it follows that the only way that $\lim_{N\to\infty} P_e^{(N)} = 0$ is that $\epsilon_{11} \to 0$ as $N \to \infty$. The conditions on R under which this happens are summarized in Lemma A2. Specifically, recalling that $\Gamma = R - (\gamma_1 + \gamma_2)$, we obtain from Lemma A2 that $P_e^{(N)} \to 0$ only if

$$
\begin{aligned}
R - (\gamma_1 + \gamma_2) &\leq\ 2n_d \\
R - (\gamma_1 + \gamma_2) &\leq\ (n_d - n_c)^+ + \max(n_d, n_c) \\
R - (\gamma_1 + \gamma_2) &\leq\ 2\max\{(n_d - n_c)^+, n_c\}.
\end{aligned}
$$

$$\tag{A24}$$
$$\tag{A25}$$
$$\tag{A26}$$

Since $\gamma_1, \gamma_2 > 0$ are arbitrary, we obtain the converse bounds (13) and (14) in Theorem 2 from (A24)–(A26) upon letting $N \to \infty$ and then $\gamma_1 \to 0$ and $\gamma_2 \to 0$.

When $p = 0$, the only positive probability is p_{00}. A necessary condition for $\lim_{N\to\infty} P_e^{(N)} = 0$ is that $\epsilon_{00} \to 0$ as $N \to \infty$. By following the same approach as for the case $p > 0$, we obtain the converse bound (12) in Theorem 2.

Appendix A.2. Converse Proof of Theorem 3

In this section, we analyze the opportunistic rate $\Delta R_1(b_1) + \Delta R_2(b_2)$, $b_i \in \{0, 1\}$ for local CSIRT and independent B_1 and B_2. Let us denote by $\hat{P}_{e_{1(b_1)}}^{(N)}$ and $\hat{P}_{e_{2(b_2)}}^{(N)}$ the error probabilities at decoders one and two, defined in (6) and (7), i.e.,

$$
\begin{aligned}
\hat{P}_{e_{1(b_1)}}^{(N)} &\triangleq\ \Pr\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))|B_1 = b_1\},\ b_1 \in \{0, 1\}, \\
\hat{P}_{e_{2(b_2)}}^{(N)} &\triangleq\ \Pr\{(\hat{W}_2, \Delta\hat{W}_2(B_2)) \neq (W_2, \Delta W_2(B_2))|B_2 = b_2\},\ b_i \in \{0, 1\}.
\end{aligned}
$$

$$\tag{A27}$$
$$\tag{A28}$$

Before we apply the Verdú-Han lemma, we have to deal with the fact that (A27) and (A28) are conditioned on two different variables but we need to analyze the probability of error jointly. To solve this problem, we expand the probability of error (A27) as

$$
\hat{P}_{e_{1(b_1)}}^{(N)} = \sum_{b_2 = 0,1} \Pr\{B_2 = b_2\} \Pr\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))|\mathbf{B} = \mathbf{b}\}.
\tag{A29}
$$

Since, by assumption, $\Pr\{B_2 = b_2\} \in (0, 1)$, it follows that

$$
\Pr\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))|B_1 = b_1\} \to 0 \text{ as } N \to \infty
$$

if, and only if,

$$\Pr\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))|\mathbf{B} = \mathbf{b}\} \to 0, \ b_2 \in \{0,1\} \text{ as } N \to \infty. \tag{A30}$$

We shall lower-bound (A29) by considering only one of the two terms in the sum. Proceeding analogously for the second user and applying the Verdú-Han lemma (Lemma A4), we obtain

$$\hat{P}_{e_{1(b_1)}}^{(N)} \geq \left(\Pr\left\{\tfrac{1}{N}i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) \leq R_1 + \Delta R_1(B_1) - \gamma_1|\mathbf{B} = \mathbf{b}\right\} - e^{-\gamma_1 N}\right)\Pr\{B_2 = b_2\}, \quad b_2 = 0,1, \tag{A31}$$

$$\hat{P}_{e_{2(b_2)}}^{(N)} \geq \left(\Pr\left\{\tfrac{1}{N}i_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b}) \leq R_2 + \Delta R_2(B_2) - \gamma_2|\mathbf{B} = \mathbf{b}\right\} - e^{-\gamma_2 N}\right)\Pr\{B_1 = b_1\}, \quad b_1 = 0,1. \tag{A32}$$

Let $\Gamma_i = R_i + \Delta R_i - \gamma_i$, $i = 1,2$ and $\Gamma = R + \Delta R_1(B_1) + \Delta R_2(B_2) - (\gamma_1 + \gamma_2)$. Then, (A31) and (A32) can be written as

$$\hat{P}_{e_{1(b_1)}}^{(N)} \geq \left(\Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N}\right)\Pr\{B_2 = b_2\}, \quad b_2 = 0,1, \tag{A33}$$

$$\hat{P}_{e_{2(b_2)}}^{(N)} \geq \left(\Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_2 N}\right)\Pr\{B_1 = b_1\}, \quad b_1 = 0,1. \tag{A34}$$

Proceeding analogously as in (A19)–(A22), and using that $\Pr\{B_i = b_i\} \geq \min\{p, 1-p\}$, we obtain

$$\begin{aligned}
\hat{P}_{e_{1(b_1)}}^{(N)} + \hat{P}_{e_{2(b_2)}}^{(N)} &\geq \left(\Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B} = \mathbf{b}\} + \Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\right)\min\{p, 1-p\} \\
&\geq \left(\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N} - e^{\gamma_2 N}\right)\min\{p, 1-p\}.
\end{aligned} \tag{A35}$$

Since $\gamma_1, \gamma_2 > 0$, the left-hand side (LHS) of (A35) only tends to zero as $N \to \infty$ if $\Pr(\mathcal{E}_{12}(\Gamma)|\mathbf{B} = \mathbf{b}) \to 0$ as $N \to \infty$. It thus follows from Lemma A2 that $\hat{P}_{e_{1(b_1)}}^{(N)} + \hat{P}_{e_{2(b_2)}}^{(N)} \to 0$ as $N \to \infty$ only if conditions (A6)–(A8) are satisfied. Letting $\gamma_1 \to 0$ and $\gamma_2 \to 0$ then gives the following constraints:

- For $\mathbf{B} = [1,1]$

$$
\begin{aligned}
R_1 + \Delta R_1(1) + R_2 + \Delta R_2(1) &\leq 2n_d \tag{A36} \\
R_1 + \Delta R_1(1) + R_2 + \Delta R_2(1) &\leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{A37} \\
R_1 + \Delta R_1(1) + R_2 + \Delta R_2(1) &\leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{A38}
\end{aligned}
$$

- For $\mathbf{B} = [0,0]$,

$$R_1 + \Delta R_1(0) + R_2 + \Delta R_2(0) \leq 2n_d. \tag{A39}$$

- For $\mathbf{B} = [0,1]$, using that $\Delta R_2(1) = 0$,

$$R_1 + \Delta R_1(0) + R_2 \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A40}$$

- For $\mathbf{B} = [1,0]$, using that $\Delta R_1(1) = 0$,

$$R_1 + R_2 + \Delta R_2(0) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A41}$$

The constraints (A39)–(A41) yield (15)–(17). This proves Theorem 3.

Appendix A.3. Achievability Proof of Theorem 3

In this section, we present the achievability bounds in Theorem 3 for the regions in which it is possible to transmit opportunistic messages, namely the VWI and WI regions. The presented bounds are valid for local CSIR and local CSIRT.

Appendix A.3.1. Very Weak Interference

Transmitter 1 (Tx$_1$) and transmitter 2 (Tx$_2$) transmit in the most significant levels a block of $n_d(1-\alpha)$ bits, and they transmit in the least significant levels a block of $n_d\alpha$ bits. The same construction is used for both transmitters. Figure A1 depicts the signal levels of the transmitted signals (normalized by n_d) as observed at receiver 1 (Rx$_1$), when it is affected by interference.
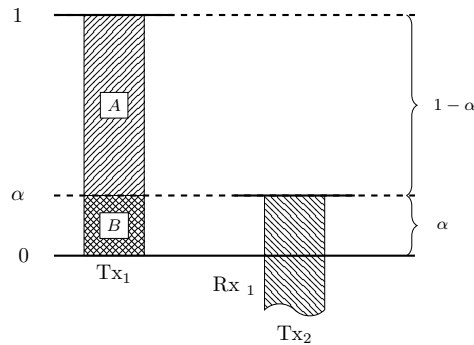
Figure A1. Normalized signal levels at Rx$_1$ for $\alpha \leq \frac{1}{2}$.

At the receiver side, we have the following procedure:

- In presence of interference: decode block $\boxed{A}$ in the desired signal which is interference free, and treat the block $\boxed{B}$ as noise. We thus obtain the individual rate

$$R_1 = (n_d - n_c)^+ \; \frac{\text{bits}}{\text{sub-channel use}}. \tag{A42}$$

- In absence of interference: decode blocks $\boxed{A}$ and $\boxed{B}$. We thus obtain the individual rate

$$R_1 + \Delta R_1(0) = n_d \; \frac{\text{bits}}{\text{sub-channel use}}. \tag{A43}$$

where $\Delta R_1(0) = n_c \frac{\text{bits}}{\text{sub-channel use}}$ corresponds to the opportunistic rate.

The bounds (A42) and (A43) coincide with the bounds for the bounds of user 2. In order to obtain the possible sum rates according to the interference states, we combine (A42) (which corresponds to $B_1 = 1$) and (A43) (which corresponds to $B_1 = 0$) to obtain the converse bounds (15)–(16).

Appendix A.3.2. Weak Interference

The symbol transmitted by Tx$_1$ (normalized by n_d) is depicted in Figure A2a. Specifically, we transmit in the most significant levels a block of $n_d(1-\alpha)$ bits. In the subsequent levels we transmit a block of $n_d(2\alpha-1)$ zeros, followed by $n_d(2-3\alpha)$ opportunistic bits. Finally, in the least significant levels, we transmit a block of $n_d(2\alpha-1)$ bits. The same construction is used for both transmitters.

Figure A2b depicts the normalized signal levels of the transmitted signals as observed by Rx$_1$. At the receiver side, we have the following procedure:

- In presence of interference: The channel pushes the interference level by $n_d - n_c$ bits. Thus, the least significant $2n_c - n_d$ bits of the desired signal (block $\boxed{A}$) align with the zeros of the interference signal and can be decoded free from interference. Since $(n_d - n_c) \leq n_c$, the most significant $n_d - n_c$ bits (block $\boxed{B}$) are also free from interference. Thus, we achieve the rate

$$\begin{aligned} R_1 &= n_d - n_c + 2n_c - n_d \\ &= n_c \; \frac{\text{bits}}{\text{sub-channel use}}. \end{aligned} \tag{A44}$$

- In absence of interference: The bits in blocks $\boxed{A}$, $\boxed{B}$, and $\boxed{D}$ can be decoded free from interference. Thus, we achieve the rate

$$
\begin{aligned}
R_1 + \Delta R_1(0) &= n_d - n_c + 2n_c - n_d + 2n_d - 3n_c \\
&= 2(n_d - n_c) \tfrac{\text{bits}}{\text{sub-channel use}}
\end{aligned}
\tag{A45}
$$

where $\Delta R_1(0) = 2n_d - 3n_c \tfrac{\text{bits}}{\text{sub-channel use}}$ corresponds to the opportunistic rate.

By symmetry, the bounds (A44) and (A45) also apply for the achievable rates of user 2. In order to obtain the possible sum rates according to the interference states, we combine (A44) (which corresponds to $B_1 = 1$) and (A45) (which corresponds to $B_1 = 0$) to obtain the achievability bounds in Theorem 3.
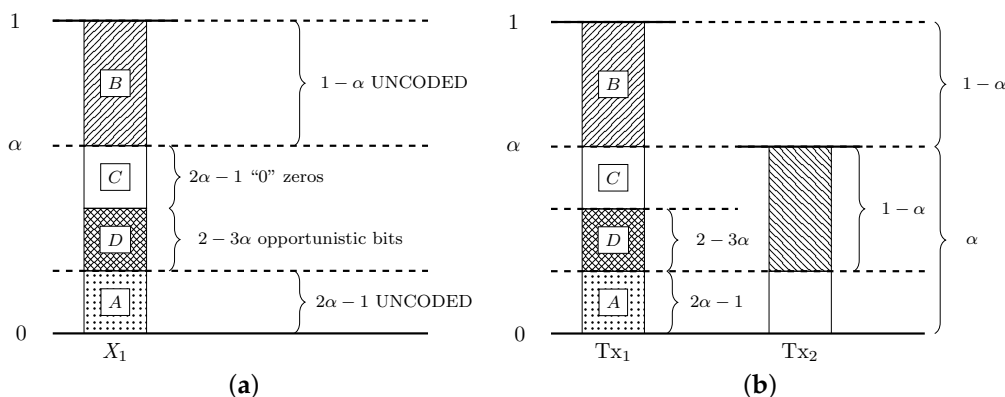
Figure A2. (**a**) Normalized transmitted symbol at Tx_1; (**b**) Normalized signal levels at Rx_1.

Appendix A.4. Converse Proof of Theorem 3 when $B_1 = B_2$

The proof of the converse bound (15) for local CSIR when $B_1 = B_2$ is similar to the proof when B_1 and B_2 are independent; see Appendix A.2. However, to prove the converse bound (16) for the case where $B_1 = B_2$ we cannot simply reproduce the steps for the independent case. The reason is that, in the correlated case, we only have the interference states $[0,0]$ and $[1,1]$, but the derivation of (16) for the independent case follows from the analysis of the states $\mathbf{B} = [0,1]$ and $\mathbf{B} = [1,0]$ (see (A40) and (A41) in Appendix A.2). To sidestep this problem, we follow a slightly different approach. Specifically, we combine the error probability of user 1 when $\mathbf{B} = [0,0]$ with that of user 2 when $\mathbf{B} = [1,1]$. This approach yields a tighter converse bound compared to the one obtained by simply considering $\mathbf{B} = [0,0]$ in both probabilities.

Consider $\hat{P}_{e_{1(b_1)}}^{(N)}$ and $\hat{P}_{e_{2(b_2)}}^{(N)}$ defined in (A27) and (A28). Applying the Verdú-Han lemma (Lemma A1) with $\Gamma_1 = R_1 + \Delta R_1(0) - \gamma_1$ and $\Gamma_2 = R_2 - \gamma_2$, and using (A29), we obtain the lower bounds

$$
\hat{P}_{e_{1(0)}}^{(N)} \geq \left(\Pr\left\{ \mathcal{E}_1(\Gamma_1) | \mathbf{B} = [0,0] \right\} - e^{-\gamma_1 N} \right) \Pr\{B_2 = 0\}
\tag{A46}
$$

$$
\hat{P}_{e_{2(1)}}^{(N)} \geq \left(\Pr\left\{ \mathcal{E}_2(\Gamma_2) | \mathbf{B} = [1,1] \right\} - e^{-\gamma_2 N} \right) \Pr\{B_1 = 1\}.
\tag{A47}
$$

Note that compared to the derivation in Section A.2, the two error events $\mathcal{E}_1(\Gamma_1)$ and $\mathcal{E}_2(\Gamma_2)$ are conditioned on different interference states. In order to derive a joint error event for $\mathcal{E}_1(\Gamma_1)$ and $\mathcal{E}_2(\Gamma_2)$, we use the next lemma.

Lemma A3. *For local CSIR, the information density* i_i, $i = 1, 2$ *depends only on* $(\mathbf{x}_i^N, \mathbf{y}_i^N)$ *and the corresponding state* b_i, *i.e.*,

$$i_1(\mathbf{x}_1^N; \mathbf{y}_1^N, [b_1, 0]) = i_1(\mathbf{x}_1^N; \mathbf{y}_1^N, [b_1, 1]) \triangleq i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, b_1) \tag{A48}$$

$$i_2(\mathbf{x}_2^N; \mathbf{y}_2^N, [0, b_2]) = i_2(\mathbf{x}_2^N; \mathbf{y}_2^N, [1, b_2]) \triangleq i_2(\mathbf{x}_2^N, \mathbf{y}_2^N, b_2). \tag{A49}$$

Proof. We prove (A48) for user 1. By the definition of the information density (A1), it follows that

$$i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, [b_1, b_2]) = \log \frac{P_{\mathbf{Y}_1^N | \mathbf{X}_1^N, \mathbf{B}}(\mathbf{y}_1^N | \mathbf{x}_1^N, [b_1, b_2])}{P_{\mathbf{Y}_1^N | \mathbf{B}}(\mathbf{y}_1^N | [b_1, b_2])} \tag{A50}$$

Evaluating i_1 for $\mathbf{B} = [0, b_2]$, $b_2 = 0, 1$ and $\mathbf{B} = [1, b_2]$, $b_2 = 0, 1$ we obtain that both cases are independent of b_2. The identity (A48) can be proven in the same way. $\square$

We next analyze the probability terms in (A46) and (A47). It follows from (A48) in Lemma A3 that $i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, [0, b_2])$ is independent of b_2. Consequently,

$$\begin{aligned} \Pr\{\mathcal{E}_1(\Gamma_1) | \mathbf{B} = [0, 0]\} &= \mathsf{E}\left[\mathbb{1}\left\{\frac{1}{N} i_1(\mathbf{X}_1^N, \mathbf{Y}_1^N, [0, 0]) \leq \Gamma_1\right\}\right] \\ &= \mathsf{E}\left[\mathbb{1}\left\{\frac{1}{N} i_1(\mathbf{X}_1^N, \mathbf{Y}_1^N, [0, 1]) \leq \Gamma_1\right\}\right] \\ &= \Pr\{\mathcal{E}_1(\Gamma_1) | \mathbf{B} = [0, 1]\}. \end{aligned} \tag{A51}$$

Analogously, using (A49) in (A47), we obtain

$$\Pr\{\mathcal{E}_2(\Gamma_2) | \mathbf{B} = [1, 1]\} = \Pr\{\mathcal{E}_2(\Gamma_2) | \mathbf{B} = [0, 1]\}. \tag{A52}$$

Adding (A46) and (A47), using (A51) and (A52), and lower-bounding $\Pr\{B_1 = 1\}$ and $\Pr\{B_2 = 0\}$ by $\min\{p, 1 - p\}$, we obtain

$$\begin{aligned} \hat{P}_{e_{1(0)}}^{(N)} + \hat{P}_{e_{2(1)}}^{(N)} &\geq \left(\Pr\{\mathcal{E}_1(\Gamma_1) | \mathbf{B} = [0, 1]\} + \Pr\{\mathcal{E}_2(\Gamma_2) | \mathbf{B} = [0, 1]\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\right) \min\{p, 1 - p\} \\ &\geq \left(\Pr\{\mathcal{E}_{12}(\Gamma) | \mathbf{B} = [0, 1]\} - e^{-\gamma_1 N} - e^{\gamma_2 N}\right) \min\{p, 1 - p\} \end{aligned} \tag{A53}$$

where $\Gamma = \Gamma_1 + \Gamma_2$. We next apply Lemma A2 with $\Gamma = R + \Delta R_1(0) + \Delta R_2(0) - (\gamma_1 + \gamma_2)$. Since $\min\{p, 1 - p\}$ is strictly positive for $0 < p < 1$, and since $-e^{-\gamma_1 N} - e^{-\gamma_2 N} \to 0$ as $N \to \infty$ for any fixed $\gamma_1, \gamma_2 > 0$, a necessary condition for (A53) going to zero is that $\Pr\{\mathcal{E}_{12}(\Gamma) | \mathbf{B} = [0, 1]\} \to 0$ as $N \to \infty$. This is the case if, and only if, (A7) in Lemma A2 is fulfilled. Since $\gamma_1, \gamma_2 > 0$ are arbitrary, we conclude the proof by letting $\gamma_1 \to 0$ and $\gamma_2 \to 0$ and using that $\Delta R_2(1) = 0$ to obtain

$$R_1 + \Delta R_1(0) + R_2 \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A54}$$

Given the symmetry of the problem, a bound on $\Delta R_2(0)$ follows by swapping the roles of users 1 and 2, yielding in this case

$$R_1 + R_2 + \Delta R_2(0) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A55}$$

Finally, combining (A54) and (A55), we obtain the bound (16) in Theorem 3 for the fully correlated scenario.

Appendix A.5. Converse Proof of Theorem 6

In this section, we analyze the opportunistic rates $\{\Delta R(\mathbf{b}), \mathbf{b} \in \{0,1\}^2\}$ for global CSIRT and independent B_1 and B_2. Let us denote by $\hat{P}_{e_{1(\mathbf{b})}}^{(N)}$ and $\hat{P}_{e_{2(\mathbf{b})}}^{(N)}$ the error probabilities at decoders 1 and 2, defined in (6) and (7), namely,

$$\hat{P}_{e_{1(\mathbf{b})}}^{(N)} \triangleq \Pr\{(\hat{W}_1, \{\Delta\hat{W}_1(\mathbf{B})\}) \neq (W_1, \{\Delta W_1(\mathbf{B})\}) | \mathbf{B} = \mathbf{b}\}, \quad \mathbf{b} \in \{0,1\}^2, \tag{A56}$$

$$\hat{P}_{e_{2(\mathbf{b})}}^{(N)} \triangleq \Pr\{(\hat{W}_2, \{\Delta\hat{W}_2(\mathbf{B})\}) \neq (W_2, \{\Delta W_2(\mathbf{B})\}) | \mathbf{B} = \mathbf{b}\}, \quad \mathbf{b} \in \{0,1\}^2. \tag{A57}$$

We shall follow analogous steps as in Section A.2 and set $\Gamma_i = R_i + \Delta R_i(\mathbf{B}) - \gamma_i$, $i = 1, 2$, and $\Gamma = R + \Delta R(\mathbf{B}) - (\gamma_1 + \gamma_2)$. Proceeding analogously as in (A19)–(A21), we obtain

$$\hat{P}_{e_{1(\mathbf{b})}}^{(N)} + \hat{P}_{e_{2(\mathbf{b})}}^{(N)} \geq \Pr\{\mathcal{E}_{12}(\Gamma) | \mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}. \tag{A58}$$

By invoking Lemma A2 for fixed (but arbitrary) $\gamma_1, \gamma_2 > 0$, and letting then $\gamma_1 \to 0$ and $\gamma_2 \to 0$, we obtain that the RHS of (A58) vanishes as $N \to \infty$ only if the following constraints are satisfied:

- For $\mathbf{B} = [1,1]$,

$$R_1 + \Delta R_1(11) + R_2 + \Delta R_2(11) \leq 2n_d \tag{A59}$$

$$R_1 + \Delta R_1(11) + R_2 + \Delta R_2(11) \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{A60}$$

$$R_1 + \Delta R_1(11) + R_2 + \Delta R_2(11) \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{A61}$$

- For $\mathbf{B} = [0,0]$,

$$R_1 + \Delta R_1(00) + R_2 + \Delta R_2(00) \leq 2n_d. \tag{A62}$$

- For $\mathbf{B} = [0,1]$,

$$R_1 + \Delta R_1(01) + R_2 + \Delta R_2(01) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A63}$$

- For $\mathbf{B} = [1,0]$,

$$R_1 + \Delta R_1(10) + R_2 + \Delta R_2(10) = (n_d - n_c)^+ + \max(n_d, n_c). \tag{A64}$$

This proves the converse bounds in Theorem 6.

Appendix A.6. Achievability Proof of Theorem 6

In this section, we present the achievability schemes for global CSIRT when B_1 and B_2 are independent. In contrast to the local CSIR/CSIRT case, we can adapt our transmission strategy to the interference states.

When $\mathbf{B} = [0,0]$, the capacity-achieving scheme consists of sending uncoded bits in all n_d level. We thus achieve the sum rate $R + \Delta R(00) = 2n_d \frac{\text{bits}}{\text{sub-channel use}}$.

When $\mathbf{B} = [0,1]$ or $\mathbf{B} = [1,0]$, the achievability schemes coincide with the schemes described in Section A.3. In this case, we can only send opportunistic messages when we have VWI or WI.

Appendix A.6.1. Very Weak Interference

Consider the achievability scheme depicted in Figure A1. By (A42) and (A43),

$$R_1 + \Delta R_1(01) = R_2 + \Delta R_2(10) = n_d \frac{\text{bits}}{\text{sub-ch.use}} \tag{A65}$$

$$R_1 + \Delta R_1(10) = R_2 + \Delta R_2(01) = n_d - n_c \frac{\text{bits}}{\text{sub-ch.use}}. \tag{A66}$$

This proves the achievability bounds in Theorem 6 for VWI.

Appendix A.6.2. Weak Interference

Consider the achievability scheme depicted in Figure A2a. By (A44) and (A45),

$$R_1 + \Delta R_1(01) = R_2 + \Delta R_2(10) \quad = \quad 2(n_d - n_c) \frac{\text{bits}}{\text{sub-ch.use}} \tag{A67}$$

$$R_1 + \Delta R_1(10) = R_2 + \Delta R_1(01) \quad = \quad n_c \frac{\text{bits}}{\text{sub-ch.use}}. \tag{A68}$$

Combining (A67) and (A68), we obtain the achievability bounds in Theorem 6 for WI.

Appendix B. Proofs for the Ergodic Case

Appendix B.1. Proof of (18) *in Theorem 4*

The bound (18) coincides with [7] (Theorem 3.1). However, [7] (Theorem 3.1) derives (18) for the considered channel model with $T = 1$ and feedback. In this section we show that (18) also holds for general T in the no-feedback case. We follow along the lines of the proof of [7] (Theorem 3.1). We begin by applying Fano's inequality to obtain

$$
\begin{aligned}
N(R_1 - \epsilon_{1K}) \leq & I(W_1; \mathbf{Y}_1^K | B_1^K) \\
= & \sum_{k=1}^{K} \left[H(\mathbf{Y}_{1,k} | \mathbf{Y}_1^{k-1}, B_1^K) - H(\mathbf{Y}_{1,k} | W_1, \mathbf{Y}_1^{k-1}, B_1^K) \right] \\
\stackrel{(a)}{=} & \sum_{k=1}^{K} \left[H(\mathbf{Y}_{1,k} | \mathbf{Y}_1^{k-1}, B_{1,k}, B_1^{k-1}, B_{1,k+1}^K) - H(B_{1,k} \mathsf{S}_{n_c} \mathbf{X}_{2,k} | \{B_{1,\ell} \mathsf{S}_{n_c} \mathbf{X}_{2,\ell}\}_{\ell=1}^{k-1}, W_1, B_1^K) \right] \\
= & \sum_{k=1}^{K} \left[(1-p) H(\mathbf{Y}_{1,k} | \mathbf{Y}_1^{k-1}, B_{1,k} = 0, B_1^{k-1}, B_{1,k+1}^K) + p H(\mathbf{Y}_{1,k} | \mathbf{Y}_1^{k-1}, B_{1,k} = 1, B_1^{k-1}, B_{1,k+1}^K) \right. \tag{A69} \\
& \quad \left. - p H(\mathsf{S}_{n_c} \mathbf{X}_{2,k} | \{B_{1,\ell} \mathsf{S}_{n_c} \mathbf{X}_{2,\ell}\}_{\ell=1}^{k-1}, W_1, B_{1,k} = 1, B_{1,k+1}^K, B_1^{k-1}) \right] \\
\stackrel{(b)}{\leq} & \sum_{k=1}^{K} \left[(1-p) H(\mathsf{S}_{n_d} \mathbf{X}_{1,k} | B_{1,k} = 0) + p H(\mathbf{Y}_{1,k} | B_{1,k} = 1) \right. \\
& \quad \left. - p H(\mathsf{S}_{n_c} \mathbf{X}_{2,k} | \{B_{1,\ell} \mathsf{S}_{n_c} \mathbf{X}_{2,\ell}\}_{\ell=1}^{k-1}, B_1^{k-1}) \right]
\end{aligned}
$$

where $\epsilon_{1K} \to 0$ as $K \to \infty$. Here, (a) follows because (W_1, B_1^K) determine $\mathbf{X}_1^K$, so we can subtract the contribution of $\mathbf{X}_1^K$ in the second entropy and by evaluating the entropy for different interference states. Step (b) follows because $(B_1^{k-1}, \mathbf{X}_2^k)$ are independent of $(B_{1,k}^K, W_1)$ (which in turn follows because $\mathbf{X}_2^K$ only depends on (B_2^K, W_2), which is independent of (B_1^K, W_1)) and because conditioning reduces entropy.

Likewise, we have

$$
\begin{aligned}
N(R_2 - \epsilon_{2K}) \leq & I(W_2; \mathbf{Y}_2^K | B_2^K) \\
\stackrel{(a)}{\leq} & I(W_2; \mathbf{Y}_1^K, \mathbf{Y}_2^K | W_1, B_1^K, B_2^K) \\
= & H(\mathbf{Y}_1^K, \mathbf{Y}_2^K | W_1, B_1^K, B_2^K) \\
= & \sum_{k=1}^{K} H(\mathbf{Y}_{1,k}, \mathbf{Y}_{2,k} | W_1, B_1^K, B_2^K, \mathbf{Y}_1^{k-1}, \mathbf{Y}_2^{k-1}) \tag{A70} \\
\stackrel{(b)}{\leq} & \sum_{k=1}^{K} H(\mathsf{S}_{n_c} \mathbf{X}_{2,k}, \mathsf{S}_{n_d} \mathbf{X}_{2,k} | W_1, B_1^K, \{B_{1,\ell} \mathsf{S}_{n_c} \mathbf{X}_{2,\ell}\}_{\ell=1}^{k-1}) \\
\stackrel{(c)}{\leq} & \sum_{k=1}^{K} \left[H(\mathsf{S}_{n_c} \mathbf{X}_{2,k} | \{B_{1,\ell} \mathsf{S}_{n_c} \mathbf{X}_{2,\ell}\}_{\ell=1}^{k-1}, B_1^{k-1}) + H(\mathsf{S}_{n_d} \mathbf{X}_{2,k} | \mathsf{S}_{n_c} \mathbf{X}_{2,k}) \right]
\end{aligned}
$$

where $\epsilon_{2K} \to 0$ as $K \to \infty$. Here, (a) follows because W_2, W_1 and B_1^K are independent. Step (b) follows because (W_1, B_1^K) determines X_1^K, so we can subtract its contribution from $(Y_{1,k}, Y_{2,k})$, because $Y_{1,k} \oplus S_{n_d} X_{1,k} = B_{1,k} S_{n_c} X_{2,k}$ has a lower entropy than $S_{n_c} X_{2,k}$, and because conditioning reduces entropy. Step (c) follows by the chain rule, and because conditioning reduces entropy.

Combining (A69) and (A70) yields

$$N(R_1 + pR_2) - N(\epsilon_{1K} + p\epsilon_{2K}) \leq \sum_{k=1}^{K} \Big[(1-p)H(S_{n_d} X_{1,k} | B_{1,k} = 0)$$
$$+ pH(Y_{1,k} | B_{1,k} = 1) + pH(S_{n_d} X_{2,k} | S_{n_c} X_{2,k}) \Big]. \tag{A71}$$

By maximizing the individual entropies in (A71) over all input distributions, dividing both sides of (A71) by $N = KT$, and by letting then K tend to infinity, we obtain that

$$R_1 + pR_2 \leq (1-p)n_d + p[(n_d - n_c)^+ + \max(n_d, n_c)]. \tag{A72}$$

By symmetry, the same bound also holds for $R_2 + pR_1$. Thus, by averaging over the two cases, it follows that (A72) is also an upper bound on $(R_1 + R_2)(1 + p)/2$. The final result (18) follows by dividing (A72) by $\frac{1+p}{2}$.

Appendix B.2. Achievability Proof of Theorem 5

In this section, we describe the achievability schemes that yield the rates presented in Theorem 5 for local CSIR. The bursty IC described in Section 2 is treated here as a set of n_d parallel sub-channels.

Appendix B.2.1. Scheme 1 (VWI; WI, MI for $0 \leq p \leq \frac{1}{2}$)

The achievability scheme is illustrated in Figure A3a. In the figure, we present the normalized received signal at Rx$_1$, i.e., we represent graphically the time-k channel output $Y_{1,k}$ given by (3), where the signal level from Tx$_1$ corresponds to $S_{n_d} X_{1,k}$ and the signal level from Tx$_2$ corresponds to $S_{n_c} X_{2,k}$, both normalized by n_d. In our scheme, the upper $n_d - n_c$ sub-channels (block $\boxed{A}$ in the figure) carry uncoded data (rate 1 bits/sub-channel use), while in the lower n_c channels (block $\boxed{B}$ in the figure) a capacity-achieving code of blocklength $N = KT$ for a *binary erasure channel (BEC)* with erasure probability p is used (with asymptotic rate $1 - p$ bits/sub-channel use) [28] (Section 7.1.5). Block $\boxed{A}$ is received free of interference and can be directly decoded at the receiver. Block $\boxed{B}$ is affected by interference with probability (w.p.) p. Since the fading state $B_{i,k}$ is known to the i-th receiver, interfered slots are treated as erasures. Consequently, when K tends to infinity, user i achieves the rate $R_i = (n_d - n_c) + (1 - p)n_c$. The sum rate R is thus given by

$$R = 2(n_d - pn_c), \quad n_d \geq n_c. \tag{A73}$$

This scheme is tight for VWI and for WI and MI when $p \leq \frac{1}{2}$.

Appendix B.2.2. Scheme 2 (WI, $\frac{1}{2} < p \leq 1$)

We next consider the achievability scheme illustrated in Figure A3b. In blocks $\boxed{A}$ and $\boxed{B}$ uncoded data is transmitted (rate 1 bits/sub-channel use), block $\boxed{C}$ carries the deterministic all-zeros sequence (rate 0 bit/sub-channel use) and in block $\boxed{D}$ a capacity-achieving code for the BEC (with asymptotic rate $1 - p$ bits/sub-channel use) is used. As in Scheme 1, blocks $\boxed{A}$ and $\boxed{B}$ can be decoded without interference, and block $\boxed{D}$ is decoded by treating interfered symbols as erasures. The rate achieved by this scheme at user i is $R_i = (n_d - n_c) + (2n_c - n_d) + (1 - p)(2n_d - 3n_c)$, so

$$R = 4(n_d - n_c) + p(6n_c - 4n_d), \quad \frac{2n_d}{3} \geq n_c \geq \frac{n_d}{2}. \tag{A74}$$
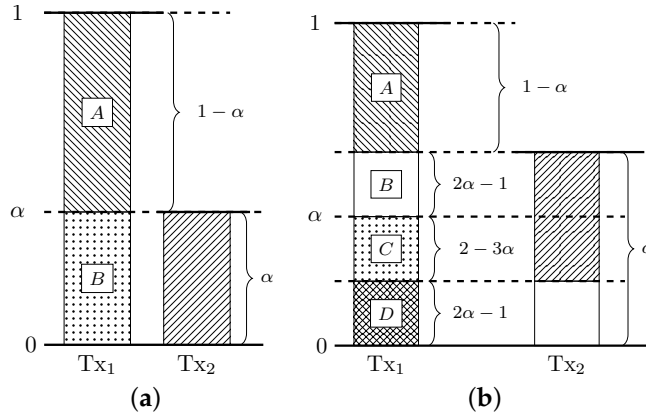
Figure A3. Normalized signal levels at Rx_1. (**a**) VWI; WI; MI, $p \leq \frac{1}{2}$; (**b**) WI, $p > \frac{1}{2}$.

Appendix B.2.3. Scheme 3 (SI, $0 \leq p \leq \frac{1}{2}$)

We use an achievability scheme similar to Scheme 1. Now, the upper $2n_d - n_c$ sub-channels carry a capacity-achieving code for a BEC with erasure probability p, and the lower $n_c - n_d$ sub-channels carry uncoded data. Consequently, when K tends to infinity, user i achieves the rate $R_i = (n_c - n_d) + (1 - p)(2n_d - n_c)$. The sum rate $R = R_1 + R_2$ is thus given by

$$R = 2(1 - 2p)n_d + 2pn_c, \quad 2n_d \geq n_c \geq n_d. \tag{A75}$$

This proves Theorem 5.

Appendix B.3. Proof of Theorem 8

In this section, we prove the converse bounds for global CSIRT and independent B_1^K and B_2^K.

Appendix B.3.1. Converse Bound (52) for Global CSIRT

By Fano's inequality, we have

$$
\begin{aligned}
N(R_1 - \epsilon_{1K}) &\leq I(W_1; Y_1^K | \mathbf{B}^K) \\
&\overset{(a)}{=} \sum_{k=1}^{K} \Big[H(Y_{1,k} | Y_1^{k-1}, \mathbf{B}^K) - H(B_{1,k} \mathsf{S}_{n_c} X_{2,k} | W_1, Y_1^{k-1}, \mathbf{B}^K) \Big] \\
&= \sum_{k=1}^{K} \Big[(1 - p) H(Y_{1,k} | Y_1^{k-1}, B_{1,k} = 0, B_1^{k-1}, B_{1,k+1}^K, B_2^K) \\
&\quad + p H(Y_{1,k} | Y_1^{k-1}, B_{1,k} = 1, B_1^{k-1}, B_{1,k+1}^K, B_2^K) \\
&\quad - p H(\mathsf{S}_{n_c} X_{2,k} | W_1, Y_1^{k-1}, B_{1,k} = 1, B_1^{k-1}, B_{1,k+1}^K, B_2^K) \Big] \\
&\leq \sum_{k=1}^{K} \Big[(1 - p) H(\mathsf{S}_{n_d} X_{1,k} | B_{1,k} = 0) + p H(Y_{1,k} | B_{1,k} = 1) \\
&\quad - p H(\mathsf{S}_{n_c} X_{2,k} | W_1, Y_1^{k-1}, B_{1,k} = 1, B_1^{k-1}, B_{1,k+1}^K, B_2^K) \Big]
\end{aligned}
\tag{A76}
$$

where $\epsilon_{1K} \to 0$ as $K \to \infty$. Here, (a) follows because $(W_1, \mathbf{B}^K)$ determines $X_{1,k}$, so we can subtract its contribution from the second entropy. Likewise,

$$
\begin{aligned}
N(R_2 - \epsilon_{2K}) &\leq I(W_2; \boldsymbol{Y}_2^K | \mathbf{B}^K) \\
&\overset{(a)}{\leq} I(W_2; \boldsymbol{Y}_1^K, \boldsymbol{Y}_2^K | W_1, \mathbf{B}^K) \\
&= \sum_{k=1}^{K} H(\boldsymbol{Y}_{1,k}, \boldsymbol{Y}_{2,k} | W_1, \boldsymbol{Y}_1^{k-1}, \boldsymbol{Y}_2^{k-1}, \mathbf{B}^K) \\
&\overset{(b)}{\leq} \sum_{k=1}^{K} H(B_{1,k} \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, \mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | W_1, \boldsymbol{Y}_1^{k-1}, \mathbf{B}^K) \\
&\overset{(c)}{\leq} \sum_{k=1}^{K} \Big[(1-p) H(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} |, B_{1,k}=0) + p H(\mathsf{S}_{n_c} \boldsymbol{X}_{2,k} | W_1, \boldsymbol{Y}_1^{k-1}, B_{1,k}=1, B_1^{k-1}, B_{1,k+1}^K, B_2^K) \\
&\quad\quad + p H(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, B_{1,k}=1) \Big]
\end{aligned}
\tag{A77}
$$

where $\epsilon_{2K} \to 0$ as $K \to \infty$. Here, step (a) follows because W_2 and (W_1, B_1^K) are independent. Step (b) follows because $(W_1, \mathbf{B}^K)$ determines $\boldsymbol{X}_{1,k}$, so we can subtract its contribution from $\boldsymbol{Y}_{1,k}$ and $\boldsymbol{Y}_{2,k}$, and because conditioning reduces entropy. Step (c) follows by evaluating the entropies for different interference states and because conditioning reduces entropy. Combining (A76) and (A77) yields

$$
\begin{aligned}
N(R_1 + R_2) - N(\epsilon_{1K} + \epsilon_{2K}) &\leq \sum_{k=1}^{K} \Big[(1-p) \big(H(\mathsf{S}_{n_d} \boldsymbol{X}_{1,k} | B_{1,k}=0) + H(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | B_{1,k}=0) \big) \\
&\quad\quad + p H(\boldsymbol{Y}_{1,k} | B_{1,k}=1) + p H(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, B_{1,k}=1) \Big] .
\end{aligned}
\tag{A78}
$$

By maximizing the entropies in (A78) over all input distributions, dividing by $N = KT$, and letting K tend to infinity, we obtain that

$$
R \leq 2(1-p) n_d + p \max(n_d, n_c) + p(n_d - n_c)^+
\tag{A79}
$$

which is (52).

Appendix B.3.2. Converse Bound (53) for Global CSIRT

Let $\mathbf{b}^K$ denote the realizations of the interference states $\mathbf{B}^K$. We label the set of time indices where the pair $(b_{1,k}, b_{2,k})$ takes the value (0,1) by A; (1,1) by B; (1,0) by C; and (0,0) by D. We denote the length of each of these states by j_A, j_B, j_C and j_D, respectively. For example,

$$
A \triangleq \{i = 1, \dots, K : \mathbf{b}_k = [1,1]\}
$$

and

$$
j_A = \sum_{k=1}^{K} \mathbb{1}\{\mathbf{B} = [1,1]\}.
$$

These states are schematically shown in Figure A4, where shaded areas correspond to $b_i = 1$.
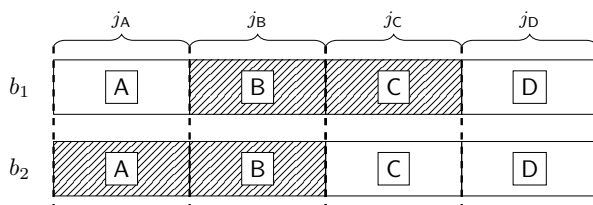
Figure A4. Possible interference states.

For global CSIRT, $(\boldsymbol{X}_1^K, \boldsymbol{X}_2^K)$ may depend on $\mathbf{B}^K = \mathbf{b}^K$. We shall denote by X_i^A, X_i^B, X_i^C and X_i^D the $\boldsymbol{X}_{1,k}$'s with indices in A, B, C and D. For example, $X_i^A = \{X_{i,k} : k \in A\}$. At time k, the interference

states $\mathbf{B}_k = \mathbf{b}_k$ can be in one of the 4 possible cases, as depicted in Figure A4. The converse bound (53) is proved as follows. We begin by applying Fano's inequality to obtain

$$
\begin{aligned}
N(R_1 + R_2) &- N(\epsilon_{1K} + \epsilon_{2K}) \\
&\leq I(W_1; Y_1^K | \mathbf{B}^K) + I(W_2; Y_2^K | \mathbf{B}^K) \\
&= \sum_{\mathbf{b} \in \{0,1\}^K} \mathcal{P}(\mathbf{B} = \mathbf{b}^K) \left[I(W_1; Y_1^K | \mathbf{B}^K = \mathbf{b}^K) + I(W_2; Y_2^K | \mathbf{B}^K = \mathbf{b}^K) \right]
\end{aligned}
\tag{A80}
$$

where $\epsilon_{1K} \to 0$ and $\epsilon_{2K} \to 0$ as $N \to \infty$. For every $\mathbf{b}^K$, we have

$$
\begin{aligned}
&I(W_1; Y_1^K | \mathbf{B}^K = \mathbf{b}^K) + I(W_2; Y_2^K | \mathbf{B}^K = \mathbf{b}^K) \\
&= H(Y_1^K | \mathbf{B}^K = \mathbf{b}^K) - H(Y_1^K | W_1, \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^K | \mathbf{B}^K = \mathbf{b}^K) - H(Y_2^K | W_2, \mathbf{B}^K = \mathbf{b}^K) \\
&\overset{(a)}{=} H(Y_1^{\mathsf{C}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{A}} Y_1^{\mathsf{B}} | Y_1^{\mathsf{C}}, \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{D}} | Y_1^{\mathsf{A}}, Y_1^{\mathsf{B}}, Y_1^{\mathsf{C}}, \mathbf{B}^K = \mathbf{b}^K) \\
&\quad - H(\mathsf{S}_{n_c} X_2^{\mathsf{B}}, \mathsf{S}_{n_c} X_2^{\mathsf{C}} | \mathbf{B}^K = \mathbf{b}^K) \\
&\quad + H(Y_2^{\mathsf{A}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{B}}, Y_2^{\mathsf{C}} | Y_2^{\mathsf{A}}, \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{D}} | Y_2^{\mathsf{A}}, Y_2^{\mathsf{B}}, Y_2^{\mathsf{C}}, \mathbf{B}^K = \mathbf{b}^K) \\
&\quad - H(\mathsf{S}_{n_c} X_1^{\mathsf{A}}, \mathsf{S}_{n_c} X_1^{\mathsf{B}} | \mathbf{B}^K = \mathbf{b}^K) \\
&\overset{(b)}{\leq} H(Y_1^{\mathsf{C}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{A}} Y_1^{\mathsf{B}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{D}} | \mathbf{B}^K = \mathbf{b}^K) - H(\mathsf{S}_{n_c} X_2^{\mathsf{B}}, \mathsf{S}_{n_c} X_2^{\mathsf{C}} | \mathbf{B}^K = \mathbf{b}^K) \\
&\quad + H(Y_2^{\mathsf{A}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{B}}, Y_2^{\mathsf{C}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{D}} | \mathbf{B}^K = \mathbf{b}^K) - H(\mathsf{S}_{n_c} X_1^{\mathsf{A}}, \mathsf{S}_{n_c} X_1^{\mathsf{B}} | \mathbf{B}^K = \mathbf{b}^K)
\end{aligned}
\tag{A81}
$$

where step (a) follows by the chain rule for entropy and because $(W_1, \mathbf{B}^K)$ determines X_1^K, so we can subtract its contribution from the second and fourth entropy. Step (b) follows because conditioning reduces entropy. We next upper-bound (A81) by combining the positive and negative entropies in areas B and C for user 1 and user 2; and areas A and B for user 2 and user 1:

$$
\begin{aligned}
&I(W_1; Y_1^K | \mathbf{B}^K = \mathbf{b}^K) + I(W_2; Y_2^K | \mathbf{B}^K = \mathbf{b}^K) \\
&\overset{(a)}{\leq} H(Y_1^{\mathsf{C}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{A}}, Y_1^{\mathsf{B}} | \mathsf{S}_{n_c} X_1^{\mathsf{A}}, \mathsf{S}_{n_c} X_1^{\mathsf{B}}, \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{D}} | \mathbf{B}^K = \mathbf{b}^K) \\
&\quad + H(Y_2^{\mathsf{A}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{B}}, Y_2^{\mathsf{C}} | \mathsf{S}_{n_c} X_2^{\mathsf{B}}, \mathsf{S}_{n_c} X_2^{\mathsf{C}}, \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{D}} | \mathbf{B}^K = \mathbf{b}^K) \\
&\leq H(Y_1^{\mathsf{C}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{A}} | \mathsf{S}_{n_c} X_1^{\mathsf{A}}, \mathbf{B}^K = \mathbf{b}^K) + H(Y_1^{\mathsf{B}} | \mathsf{S}_{n_c} X_1^{\mathsf{B}}, \mathbf{B}^K = \mathbf{b}^K) \\
&\quad + H(Y_1^{\mathsf{D}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{A}} | \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{B}} | \mathsf{S}_{n_c} X_2^{\mathsf{B}}, \mathbf{B}^K = \mathbf{b}^K) \\
&\quad + H(Y_2^{\mathsf{C}} | \mathsf{S}_{n_c} X_2^{\mathsf{C}}, \mathbf{B}^K = \mathbf{b}^K) + H(Y_2^{\mathsf{D}} | \mathbf{B}^K = \mathbf{b}^K)
\end{aligned}
\tag{A82}
$$

where step (a) follows because $H(F) - H(G) \leq H(F|G)$ for any random variables F and G. By maximizing the entropies in (A82) over all input distributions, we obtain

$$
\begin{aligned}
I(W_1; Y_1^K | \mathbf{B}^K = \mathbf{b}^K) + I(W_2; Y_2^K \mathbf{B}^K = \mathbf{b}^K) \leq\ & j_{\mathsf{A}} T[(n_d - n_c)^+ + \max(n_d, n_c)] \\
& + 2 j_{\mathsf{B}} T \max\{(n_d - n_c)^+, n_d\} \\
& + j_{\mathsf{C}} T[(n_d - n_c)^+ + \max(n_d, n_c)] + 2 j_{\mathsf{D}} T(n_d).
\end{aligned}
\tag{A83}
$$

By dividing (A83) by $N = KT$, and taking the limit as $K \to \infty$, we obtain

$$
\begin{aligned}
R_1 + R_2 \overset{(a)}{\leq}\ & \lim_{K \to \infty} \frac{1}{KT} \sum_{\mathbf{b}^K} \mathcal{P}\{\mathbf{B}^K = \mathbf{b}^K\} \left[I(W_1; Y_1^K | \mathbf{B}^K = \mathbf{b}^K) + I(W_2; Y_2^K | \mathbf{B}^K = \mathbf{b}^K) \right] \\
=\ & \lim_{K \to \infty} \frac{1}{K} \left[\mathsf{E} \left[j_{\mathsf{A}}[(n_c - n_d)^+ + \max(n_d, n_c)] + 2 j_{\mathsf{B}}[\max\{(n_d - n_c)^+, n_c\}] \right] \right. \\
& \left. + \mathsf{E}\left[j_{\mathsf{C}}[(n_d - n_c)^+ + \max(n_d, n_c)] + 2 j_{\mathsf{D}} n_d \right] \right]
\end{aligned}
\tag{A84}
$$

where (a) follows because $(\epsilon_{1K} + \epsilon_{2K}) \to 0$ as $K \to \infty$. Next, we apply the *dominated convergence theorem (DCT)* [29] (Section 1.34) to interchange limit and expectation. By the law of large numbers, we have that $\frac{j_A}{K} \to p(1-p)$, $\frac{j_B}{K} \to p^2$, $\frac{j_C}{K} \to p(1-p)$, and $\frac{j_B}{K} \to (1-p)^2$ almost surely as $K \to \infty$. By replacing these probabilities in (A84), we thus obtain

$$R \leq 2p(1-p)[(n_d - n_c)^+ + \max(n_d, n_c)] + 2p^2 \max\{(n_d - n_c)^+, n_c\} + 2(1-p)^2 n_d. \tag{A85}$$

This yields (53).

Appendix B.4. Proof of Theorem 9

In this section, we present the achievability schemes for global CSIRT and independent B_1^K and B_2^K. Let $\mathbf{b}^K$ denote the realizations of the interference states $\mathbf{B}^K$, and define $j_{\min} \triangleq \min(j_A, j_B, j_C)$. Consider the following achievable schemes.

Appendix B.4.1. Scheme 1 (MI, $0 \leq p \leq 1$)

Both transmitters employ uncoded transmission in the first $j_{\min}$ indices of regions A and C, respectively, and in the whole region D. Tx_1 copies the first $j_{\min}$ indices of region A in region B, while Tx_2 copies the first $j_{\min}$ indices of region C in B, aligned with those of user 1. The remaining indices are treated as a non-bursty IC attaining rate $r_{ic} = n_d - \frac{n_c}{2}$ [20].

To illustrate the decoding process, Figure A5 shows the different normalized signals at the Rx_1 when $j_A = j_B = j_C = j_D = 1$. Tx_1 transmits the signals $\boxed{1}$, $\boxed{3}$, and $\boxed{4}$, in channel state A and B, C, and D, respectively. Similarly, Tx_2 transmits the signal $\boxed{2}$ in states B and C. Rx_1 has access to a clean copy of signal $\boxed{1}$ in region A, which can then be subtracted in state B to recover the interfering signal $\boxed{2}$. Since Tx_2 transmits the same signal in state C, the interference can then be canceled. Hence, signals $\boxed{3}$ and $\boxed{4}$ are recovered. For a given interference state and general A and B, C, and D, the rate attained by user i with this scheme is

$$R_i(\mathbf{b}^K) = n_d \frac{2j_{\min}}{K} + n_d \frac{j_D}{K} + r_{ic} \frac{j_A + j_B + j_C - 3j_{\min}}{K}. \tag{A86}$$

Figure A5. Normalized by n_d signal levels at Rx_1 for MI and $j_A = j_B = j_C = j_D$.

Averaging (A86) over $\mathbf{B}^K$, and letting $K \to \infty$, we obtain for the sum rate

$$
\begin{aligned}
R &= \lim_{K \to \infty} 2\mathbb{E}\left[n_d \frac{2j_{\min}}{K} + n_d \frac{j_D}{K} + r_{ic} \frac{j_A + j_B + j_C - 3j_{\min}}{K}\right] \\
&= 4n_d p_{\min} + 2n_d(1-p)^2 + (2n_d - n_c)(2p - p^2 - 3p_{\min})
\end{aligned}
\tag{A87}
$$

where we changed the order of limit and expectation by appealing to the DCT, and used that, by the law of large numbers, $\frac{j_A}{K} \to p(1-p)$, $\frac{j_B}{K} \to p^2$, $\frac{j_C}{K} \to p(1-p)$ and $\frac{j_D}{K} \to (1-p)^2$ almost surely as $K \to \infty$.

Appendix B.4.2. Scheme 2 (SI, $0 \le p \le 1$)

Both transmitters employ uncoded transmission in the first $j_{\min}$ indices of states A and C. Tx_1 copies the lowest $2n_d - n_c$ sub-channels of the first $j_{\min}$ indices of region A into the highest $2n_d - n_c$ sub-channels and uses uncoded transmission in the lowest $n_c - n_d$ sub-channels of the corresponding sub-region in B. Tx_2 proceeds analogously but from region C to B. Both transmitters employ uncoded transmission in region D and treat the remaining indices as a non-bursty IC [20] with rate $\frac{n_c}{2}$.

To illustrate the decoding process, Figure A6 shows the different normalized signals at the Rx_1 when $j_A = j_B = j_C = j_D = 1$. Tx_1 transmits the signals ($\boxed{1}$,$\boxed{2}$), ($\boxed{1}$,$\boxed{3}$), $\boxed{5}$ and $\boxed{6}$ in channel state A and B, C, and D, respectively. Similarly, Tx_2 transmits the signal ($\boxed{4}$,$\boxed{7}$) and ($\boxed{4}$,$\boxed{8}$) in states B and C, respectively. Rx_1 has access to a clean copy of signals $\boxed{1}$ and $\boxed{2}$ in region A, signal $\boxed{1}$ can then be subtracted in state B to recover the interfering signals $\boxed{4}$ and $\boxed{7}$. In state B, Rx_1 has access to signal $\boxed{3}$. Since Tx_2 transmits signal $\boxed{4}$ in state C, the interference can then be canceled. Hence, signal $\boxed{5}$ can be recovered. Finally, signal $\boxed{6}$ is recovered without interference. For a given interference state, and general j_A, j_B, j_C, j_D, the rate attained by user i with this scheme is

$$R_i(\mathbf{b}^K) = (n_d + n_c)\tfrac{2j_{\min}}{K} + n_d\tfrac{j_D}{K} + r_{\text{ic}}\tfrac{j_A + j_B + j_C - 3j_{\min}}{K}. \tag{A88}$$
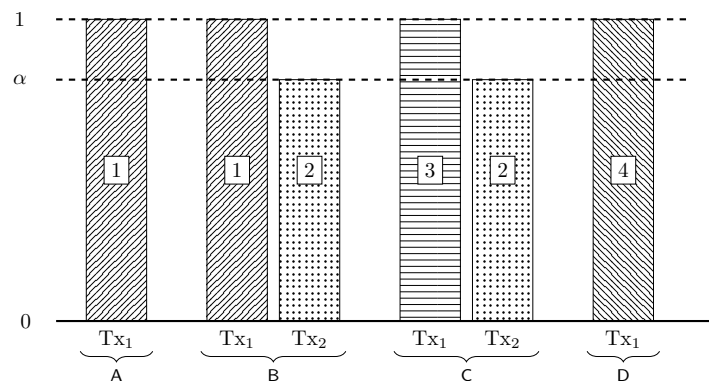
Averaging (A88) over $\mathbf{B}^K$, and letting $K \to \infty$, we obtain for the sum rate

$$R = \lim_{K\to\infty} 2\mathsf{E}\left[(n_d + n_c)\tfrac{2j_{\min}}{K} + n_d\tfrac{j_D}{K} + r_{\text{ic}}\tfrac{j_A + j_B + j_C - 3j_{\min}}{K}\right]$$
$$= 2(n_d + n_c)p_{\min} + 2n_d(1 - p)^2 + n_c(2p - p^2 - 3p_{\min}). \tag{A89}$$

where we changed the order of limit and expectation by appealing to the DCT, and used that, by the law of large numbers, $\frac{j_A}{K} \to p(1-p)$, $\frac{j_B}{K} \to p^2$, $\frac{j_C}{K} \to p(1-p)$ and $\frac{j_D}{K} \to (1-p)^2$ almost surely as $K \to \infty$.
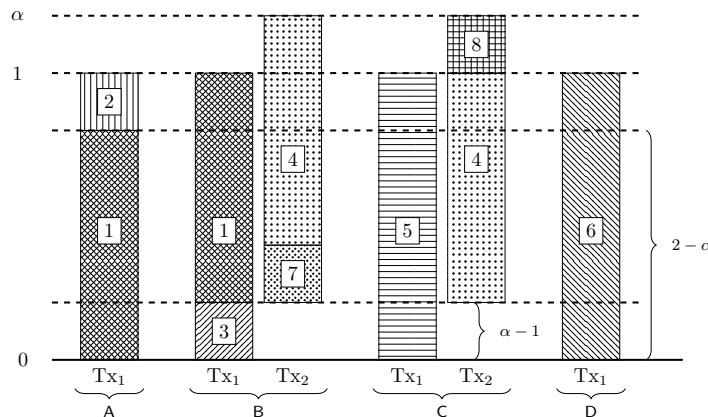
Figure A6. Normalized by n_d signal levels at Rx_1 for SI.

Appendix B.5. Proof of Theorem 10

The converse bound (59) for global CSIRT follows similar steps as in Appendix B.3.1 but considering $B_1^K = B_2^K = B^K$. We next present the converse bound (60) for global CSIRT when $B_1^K = B_2^K$. This bound follows by giving the extra information $(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K)$ to Rx_1. By Fano's inequality, we have

$$
\begin{aligned}
N(R_1 - \epsilon_{1K}) &\leq I(W_1; \boldsymbol{Y}_1^K | B^K) \\
&\leq I(W_1; \boldsymbol{Y}_1^K, B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) \\
&= I(W_1; B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) + I(W_1; \boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) \\
&= H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) + H(\boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) - H(\boldsymbol{Y}_1^K | W_1, B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) \\
&= H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) + H(\boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) - H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K | B^K)
\end{aligned}
\tag{A90}
$$

where $\epsilon_{1K} \to 0$ as $K \to \infty$. Analogously, by giving the extra information $(B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K)$ to Rx$_2$, we obtain

$$
N(R_2 - \epsilon_{2K}) \leq H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K | B^K) + H(\boldsymbol{Y}_2^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K, B^K) - H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K)
\tag{A91}
$$

where $\epsilon_{2K} \to 0$ as $K \to \infty$. Thus, (A90) and (A91) yield

$$
\begin{aligned}
N(R_1 &+ R_2) - N(\epsilon_{1K} + \epsilon_{2K}) \\
&\leq H(\boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) + H(\boldsymbol{Y}_2^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K, B^K) \\
&= \sum_{k=1}^{K} \left[H(\boldsymbol{Y}_{1,k} | \boldsymbol{Y}_1^{k-1}, B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) + H(\boldsymbol{Y}_{2,k} | \boldsymbol{Y}_2^{k-1}, B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K, B^K) \right] \\
&\leq \sum_{k=1}^{K} \left[H(\boldsymbol{Y}_{1,k} | B_k \mathsf{S}_{n_c} \boldsymbol{X}_{1,k}, B_k) + H(\boldsymbol{Y}_{2,k} | B_k \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, B_k) \right] \\
&\leq \sum_{k=1}^{K} \Big[(1-p) \left(H(\mathsf{S}_{n_d} \boldsymbol{X}_{1,k} | B_k = 0) + H(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | B_k = 0) \right) \\
&\qquad\quad + p(H(\boldsymbol{Y}_{1,k} | \mathsf{S}_{n_c} \boldsymbol{X}_{1,k}, B_k = 1) + H(\boldsymbol{Y}_{2,k} | \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, B_k = 1)) \Big]
\end{aligned}
\tag{A92}
$$

where we have used that conditioning reduces entropy. By maximizing the entropies in (A92) over all input distributions, dividing by $N = KT$, and letting K tend to infinity, we obtain that

$$
R \leq 2(1-p)n_d + 2p \max\{(n_d - n_c)^+, n_c\}.
\tag{A93}
$$

This proves (60).

Appendix C. Proof of Lemma A2

In this section, we prove the Lemma A2. To this end, we first introduce definitions and properties that will be used in the proof of the lemma.

Definition A1 (Sup-entropy rate). *The sup-entropy rate $\overline{H}(Y)$ is defined as the limsup in probability of $\frac{1}{N} \log \frac{1}{P_{Y^N}(Y^N)}$. Analogously, the conditional sup-entropy rate $\overline{H}(Y|X)$ is the limsup in probability (according to $\{P_{X^N Y^N}\}$) of $\frac{1}{N} \log \frac{1}{P_{Y^N|X^N}(Y^N|X^N)}$.*

Lemma A4 (Sup-entropy rate properties). *Suppose (X,Y) takes values in $(\mathcal{X}, \mathcal{Y})$. The sup-entropy rate has the following properties:*

$$
\overline{H}(Y|X) < \overline{H}(Y)
\tag{A94}
$$

$$
0 \leq \overline{H}(Y) < \log |\mathcal{Y}|
\tag{A95}
$$

where $|\mathcal{Y}|$ denotes the cardinality of Y.

Proof. Property (A94) follows directly from properties (c) and (d) of [27] (Theorem 8). Property (A95) is equal to property e) in [27] (Theorem 8). □

We recall the information densities $i_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b})$ and $i_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b})$ defined in (A1) and (A2), respectively. By decomposing the logarithms and applying the Bayes rule to both probability terms, we obtain

$$
\begin{aligned}
i_i(\mathbf{x}_i^N, \mathbf{y}_i^N, \mathbf{b}) &= \log P_{\mathbf{Y}_i^N|\mathbf{X}_i^N,\mathbf{B}}(\mathbf{y}_i^N|\mathbf{x}_i^N,\mathbf{b}) - \log P_{\mathbf{Y}_i^N|\mathbf{B}}(\mathbf{y}_i^N|\mathbf{b}) \\
&= \log P_{\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}}(\mathbf{x}_i^N|\mathbf{y}_i^N,\mathbf{b}) - \log P_{\mathbf{X}_i^N|\mathbf{B}}(\mathbf{x}_i^N|\mathbf{b}).
\end{aligned}
\tag{A96}
$$

To shorten notation, we shall omit the arguments and write $i_i \triangleq i_i(\mathbf{x}_i^N, \mathbf{y}_i^N, \mathbf{b})$, $i = 1, 2$ wherever the arguments are clear from the context.

Recall the error events $\mathcal{E}_i(\Gamma_i) \triangleq \left\{ \frac{1}{n} i_i \leq \Gamma_i \right\}$, $i = 1, 2$, and $\mathcal{E}_{12}(\Gamma) \triangleq \left\{ \frac{1}{n} i_1 + \frac{1}{n} i_2 \leq \Gamma \right\}$, with $\Gamma = \Gamma_1 + \Gamma_2$, as defined in (A3) and (A4), respectively. We first note that

$$
\begin{aligned}
\mathcal{E}_1 \cap \mathcal{E}_2 &\subseteq \mathcal{E}_{12} \tag{A97} \\
\mathcal{E}_1 \cap \mathcal{E}_2 &= \mathcal{E}_1 \setminus \{\mathcal{E}_1 \cap \mathcal{E}_2^c\} \supseteq \mathcal{E}_1 \setminus \{\mathcal{E}_2^c\} \tag{A98}
\end{aligned}
$$

where (A97) follows because the conditions $\frac{1}{N} i_1 \leq \Gamma_1$ and $\frac{1}{N} i_2 \leq \Gamma_2$ imply that $\frac{1}{N}(i_1 + i_2) \leq \Gamma_1 + \Gamma_2$. Then, (A98) follows by applying basic set operations. Using (A97) and (A98), and computing the probability of the corresponding events, we obtain

$$
\Pr\{\mathcal{E}_{12}\} \geq \Pr\{\mathcal{E}_1\} - \Pr\{\mathcal{E}_2^c\}.
\tag{A99}
$$

For clarity of exposition, we define

$$
\epsilon_{\mathbf{b}} \triangleq \Pr\left\{ \frac{1}{N}(i_1 + i_2) \leq \Gamma \,\Big|\, \mathbf{B} = \mathbf{b} \right\}
\tag{A100}
$$

and analyze the necessary conditions on Γ such that $\epsilon_{\mathbf{b}} \to 0$ as $N \to \infty$. We next consider separately the four possible realizations of $\mathbf{B} = \mathbf{b}$.

Appendix C.1. Case $\mathbf{B} = [0, 0]$

When $\mathbf{B} = [0, 0]$, the channel corresponds to two parallel channels with no interference links. Then, the underlying distribution of the probability (A100) is

$$
\begin{aligned}
&P_{\mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{Y}_1^N, \mathbf{Y}_2^N|\mathbf{B}}(\mathbf{x}_1^N, \mathbf{x}_2^N, \mathbf{y}_1^N, \mathbf{y}_2^N|\mathbf{b}) \\
&\qquad = P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b}) P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b}) \mathbb{1}\{\mathbf{y}_1^N = \mathsf{S}_{n_d} \mathbf{x}_1^N\} \mathbb{1}\{\mathbf{y}_2^N = \mathsf{S}_{n_d} \mathbf{x}_2^N\}
\end{aligned}
\tag{A101}
$$

as the outputs $\mathbf{y}_1^N$ and $\mathbf{y}_2^N$ must coincide with the corresponding inputs according to the deterministic model. To prove the constraint (A6), we use (A96) in (A100) to obtain

$$
\epsilon_{00} = \Pr\left\{ -\frac{1}{N} \log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N} \log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \leq \Gamma \,\Big|\, \mathbf{B} = [0, 0] \right\}
\tag{A102}
$$

where we used that, according to (A101), $\log P_{\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}}(\mathbf{X}_i^N|\mathbf{Y}_i^N, \mathbf{B}) = 0$ w.p. 1, for $i = 1, 2$.

We consider now the conditional sup-entropy rates $\overline{H}(\mathbf{X}_i^N|\mathbf{B})$, $i = 1, 2$. According to (A95) in Lemma A4, we have that $\overline{H}(\mathbf{X}_i^N|\mathbf{B}) < n_d$, $i = 1, 2$. With these considerations, if we set $\Gamma = 2n_d + 2\delta$ for some arbitrary $\delta > 0$ in (A102), we obtain

$$
\begin{aligned}
\epsilon_{00} \geq &\Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B} \leq 2n_d + 2\delta \,\Big|\, \mathbf{B} = [0,0] \right\} \\
\geq &\Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) < \overline{H}(\mathbf{X}_1^N|\mathbf{B}) + \overline{H}(\mathbf{X}_2^N|\mathbf{B}) + 2\delta \,\Big|\, \mathbf{B} = [0,0] \right\} \\
\geq &\Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) < \overline{H}(\mathbf{X}_1^N|\mathbf{B}) + \delta \,\Big|\, \mathbf{B} = [0,0] \right\} \\
&- \Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \geq \overline{H}(\mathbf{X}_2^N|\mathbf{B}) + \delta \,\Big|\, \mathbf{B} = [0,0] \right\}
\end{aligned}
\tag{A103}
$$

where the last step follows from (A99).

Recalling the definitions of the conditional sup-entropy rates $\overline{H}(\mathbf{X}_i^N|\mathbf{B})$ we have that, for any $\delta > 0$,

$$
\lim_{N\to\infty} \Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_i^N|\mathbf{B}}(\mathbf{X}_i^N|\mathbf{B}) \geq \overline{H}(\mathbf{X}_i^N|\mathbf{B}) + \delta \,\Big|\, \mathbf{B} = [0,0] \right\} = 0, \quad i = 1, 2.
\tag{A104}
$$

This implies that the first probability on the RHS of (A103) tends to 1 as $N \to \infty$, and the second probability on the RHS of (A103) tends to 0 as $N \to \infty$. We conclude that for any $\Gamma > 2n_d$ the lower bound in (A103) tends to 1 as $N \to \infty$. Thus, $\epsilon_{00} \to 0$ as $N \to \infty$ only if $\Gamma \leq 2n_d$.

Appendix C.2. Case $\mathbf{B} = [0,1]$

When $\mathbf{B} = [0,1]$, the channel corresponds to a two-user IC where only one of the transmitters interferes its non-intended receiver. In this case, the underlying distribution in (A100) is given by

$$
\begin{aligned}
P_{\mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{Y}_1^N, \mathbf{Y}_2^N|\mathbf{B}}&(\mathbf{x}_1^N, \mathbf{x}_2^N, \mathbf{y}_1^N, \mathbf{y}_2^N|\mathbf{b}) \\
&= P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b}) P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b}) \mathbb{1}\{\mathbf{y}_1^N = \mathsf{S}_{n_d}\mathbf{x}_1^N\} \mathbb{1}\{\mathbf{y}_2^N = \mathsf{S}_{n_d}\mathbf{x}_2^N \oplus \mathsf{S}_{n_c}\mathbf{x}_1^N\}
\end{aligned}
\tag{A105}
$$

We next prove the constraints (A6) and (A7) in Lemma A2.

Appendix C.2.1. Proof of Constraint (A6)

We lower-bound the probability ϵ_{01} by that of 2 parallel channels and follow the steps in Appendix C.1. Indeed, by using (A96) in (A100) and lower-bounding $\log P_{\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}}(\mathbf{X}_i^N|\mathbf{Y}_i^N, \mathbf{B}) \leq 0$, $i = 1, 2$, we obtain that

$$
\epsilon_{01} \geq \Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \leq \Gamma \,\Big|\, \mathbf{B} = [0,1] \right\}.
\tag{A106}
$$

The RHS of (A106) coincides with (A102) conditioned in $\mathbf{B} = [0,1]$. The proof then follows the one in Appendix C.1, with the probabilities and sup-entropy rates conditioned on $\mathbf{B} = [0,1]$ instead of $\mathbf{B} = [0,0]$.

Appendix C.2.2. Proof of Constraint (A7)

According to (A105), the following identities hold w.p. 1:

(i1) $\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N = \mathsf{S}_{n_c}\mathbf{X}_1^N$

(i2) $P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{X}_2^N, [0,1]) = P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N|\mathbf{B} = [0,1])$

(i3) $P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_1^N, [0,1]) = 1$

Using (A96) in (A100) and the identities (i1)–(i3), we obtain

$$
\begin{aligned}
\epsilon_{01} = \Pr\Bigg\{ & \frac{1}{N}\log P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_1^N|\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{B}) \\
& + \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \le \Gamma \,\Big|\, \mathbf{B}=[0,1] \Bigg\} \\
= \Pr\Bigg\{ & -\frac{1}{N}\log P_{\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}) \\
& + \frac{1}{N}\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \le \Gamma \,\Big|\, \mathbf{B}=[0,1] \Bigg\}
\end{aligned}
\tag{A107}
$$

We next define $\tilde{\mathsf{L}}_d \triangleq \mathsf{L}_d \mathsf{S}_{n_d}$ and apply the chain rule of probability to obtain

$$
\begin{aligned}
& \log P_{\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_d}\mathbf{x}_1^N|\mathbf{b}) \\
& = \log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_1^N|\mathbf{b}) + \log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{x}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b}).
\end{aligned}
\tag{A108}
$$

Using (A108) in (A107) and canceling the term $\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B})$, we obtain

$$
\begin{aligned}
\epsilon_{01} = \Pr\Bigg\{ & -\frac{1}{N}\log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}) \\
& -\frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \le \Gamma \,\Big|\, \mathbf{B}=[0,1] \Bigg\}.
\end{aligned}
\tag{A109}
$$

Consider the sup-entropy rates $\overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big)$ and $\overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big)$. By (A94) and (A95) in Lemma A4, we have that

$$
\overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big) \le \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathbf{B}\big) \quad < \quad (n_d-n_c)^+ \tag{A110}
$$
$$
\overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big) \quad < \quad \max(n_d,n_c). \tag{A111}
$$

Let $\Gamma = (n_d-n_c)^+ + \max(n_d,n_c) + 2\delta$ for some arbitrary $\delta > 0$. It follows that $\Gamma \ge \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big) + \overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big) + 2\delta$, so (A109) can be lower-bounded as

$$
\begin{aligned}
\epsilon_{01} \ge \Pr\Bigg\{ & -\frac{1}{N}\log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \\
& < \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big) + \overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big) + 2\delta \,\Big|\, \mathbf{B}=[0,1] \Bigg\} \\
\ge \Pr\Bigg\{ & -\frac{1}{N}\log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}) \\
& < \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big) + \delta \,\Big|\, \mathbf{B}=[0,1] \Bigg\} \\
& - \Pr\Bigg\{ -\frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \ge \overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big) + \delta \,\Big|\, \mathbf{B}=[0,1] \Bigg\}
\end{aligned}
\tag{A112}
$$

where the second step follows from (A99). By the definition of the conditional sup-entropy rate, it follows that the first probability on the RHS of (A112) tends to 1 as $N \to \infty$, and the second probability on the RHS of (A112) tends to 0 as $N \to \infty$. This implies that $\epsilon_{01} \to 0$ as $N \to \infty$ only if $\Gamma \le (n_d-n_c)^+ + \max(n_d,n_c)$ and proves conditions (A6) and (A7) in Lemma A2.

Remark A1. *Given the symmetry of the problem, the constraints (A6) and (A7) for $\mathbf{B}=[1,0]$ are proven by swapping the roles of users 1 and 2, and following the same steps as for $\mathbf{B}=[0,1]$.*

Appendix C.3. Case $\mathbf{B} = [1, 1]$

This scenario corresponds to a non-bursty IC. The underlying distribution in (A100) is given by

$$
\begin{aligned}
&P_{\mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{Y}_1^N, \mathbf{Y}_2^N | \mathbf{B}}(\mathbf{x}_1^N, \mathbf{x}_2^N, \mathbf{y}_1^N, \mathbf{y}_2^N | \mathbf{b}) \\
&= P_{\mathbf{X}_1^N | \mathbf{B}}(\mathbf{x}_1^N | \mathbf{b}) P_{\mathbf{X}_2^N | \mathbf{B}}(\mathbf{x}_2^N | \mathbf{b}) \mathbb{1}\{\mathbf{y}_1^N = \mathsf{S}_{n_d}\mathbf{x}_1^N \oplus \mathsf{S}_{n_c}\mathbf{x}_2^N\} \mathbb{1}\{\mathbf{y}_2^N = \mathsf{S}_{n_d}\mathbf{x}_2^N \oplus \mathsf{S}_{n_c}\mathbf{x}_1^N\}
\end{aligned}
\tag{A113}
$$

where the last step follows from the deterministic model since, for given $\mathbf{x}_1^N$ and $\mathbf{x}_2^N$, the outputs $\mathbf{y}_1^N$ and $\mathbf{y}_2^N$ are given by the equations appearing in the corresponding indicator functions. We next obtain the constraints (A6)–(A8) in Lemma A2.

Appendix C.3.1. Proof of Constraint (A6)

To prove this constraint, we lower-bound the probability ϵ_{11} by that of 2 parallel channels. Indeed, using (A96) in (A100), we obtain that

$$
\begin{aligned}
\epsilon_{11} &= \Pr\Bigg\{ \frac{1}{N}\Bigg(\log P_{\mathbf{X}_1^N | \mathbf{Y}_1^N, \mathbf{B}}(\mathbf{X}_1^N | \mathbf{Y}_1^N, \mathbf{B}) - \log P_{\mathbf{X}_1^N | \mathbf{B}}(\mathbf{X}_1^N | \mathbf{B}) \\
&\qquad\qquad + \log P_{\mathbf{X}_2^N | \mathbf{Y}_2^N, \mathbf{B}}(\mathbf{X}_2^N | \mathbf{Y}_2^N, \mathbf{B}) - \log P_{\mathbf{X}_2^N | \mathbf{B}}(\mathbf{X}_2^N | \mathbf{B})\Bigg) \le \Gamma \,\Big|\, \mathbf{B} = [1,1]\Bigg\} \\
&\ge \Pr\Bigg\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N | \mathbf{B}}(\mathbf{X}_1^N | \mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N | \mathbf{B}}(\mathbf{X}_2^N | \mathbf{B}) \le \Gamma \,\Big|\, \mathbf{B} = [1,1]\Bigg\}
\end{aligned}
\tag{A114}
$$

where the inequality follows because $\log P_{\mathbf{X}_i^N | \mathbf{Y}_i^N, \mathbf{B}}(\mathbf{X}_i^N | \mathbf{Y}_i^N, \mathbf{B}) \le 0$, $i = 1, 2$. As this expression coincides with (A102) conditioned on $\mathbf{B} = [1, 1]$, the proof then follows the one in Appendix C.1, with the probabilities and sup-entropy rates conditioned on $\mathbf{B} = [1, 1]$ instead of $\mathbf{B} = [0, 0]$.

Appendix C.3.2. Proof of Constraint (A7)

We next lower-bound the probability ϵ_{11} by that of an interference channel, in which only one of the transmitters interferes its non-intended receiver. Using the information densities i_1 and i_2 in (A96), we have that

$$
\begin{aligned}
i_1 + i_2 &= \log P_{\mathbf{Y}_1^N | \mathbf{X}_1^N, \mathbf{B}}(\mathbf{y}_1^N | \mathbf{x}_1^N, \mathbf{b}) - \log P_{\mathbf{Y}_1^N | \mathbf{B}}(\mathbf{y}_1^N | \mathbf{b}) + \log P_{\mathbf{Y}_2^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{y}_2^N | \mathbf{x}_2^N, \mathbf{b}) - \log P_{\mathbf{Y}_2^N | \mathbf{B}}(\mathbf{y}_2^N | \mathbf{b}) \\
&= \log P_{\mathbf{Y}_1^N | \mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{B}}(\mathbf{y}_1^N | \mathbf{x}_1^N, \mathbf{x}_2^N, \mathbf{b}) - \log P_{\mathbf{Y}_1^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{y}_1^N | \mathbf{x}_2^N, \mathbf{b}) \\
&\quad + \log P_{\mathbf{Y}_2^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{y}_2^N | \mathbf{x}_2^N, \mathbf{b}) - \log P_{\mathbf{Y}_2^N | \mathbf{B}}(\mathbf{y}_2^N | \mathbf{b}) - \log \frac{P_{\mathbf{X}_1^N | \mathbf{Y}_1^N, \mathbf{X}_2^N, \mathbf{B}}(\mathbf{x}_1^N | \mathbf{y}_1^N, \mathbf{x}_2^N, \mathbf{b})}{P_{\mathbf{X}_1^N | \mathbf{Y}_1^N, \mathbf{B}}(\mathbf{x}_1^N | \mathbf{y}_1^N, \mathbf{b})}
\end{aligned}
\tag{A115}
$$

where the second step follows from adding and subtracting

$$
\frac{1}{N}\log \frac{P_{\mathbf{Y}_1^N | \mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{B}}(\mathbf{y}_1^N | \mathbf{x}_1^N, \mathbf{x}_2^N, \mathbf{b})}{P_{\mathbf{Y}_1^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{y}_1^N | \mathbf{x}_2^N, \mathbf{b})}
$$

and simplifying the resulting terms via the Bayes rule and using that $P_{\mathbf{X}_1^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{x}_1^N | \mathbf{x}_2^N, \mathbf{b}) = P_{\mathbf{X}_1^N | \mathbf{B}}(\mathbf{x}_1^N | \mathbf{b})$ since $\mathbf{X}_1^N$ and $\mathbf{X}_2^N$ are independent conditioned on $\mathbf{B}$.

According to the underlying distribution (A113), the following identities hold w.p. 1:

(i1) $\mathbf{Y}_1^N \oplus \mathsf{S}_{n_c}\mathbf{X}_2^N = \mathsf{S}_{n_d}\mathbf{X}_1^N$

(i2) $\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N = \mathsf{S}_{n_c}\mathbf{X}_1^N$

(i3) $P_{\mathbf{Y}_1^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{Y}_1^N | \mathbf{X}_2^N, \mathbf{B} = [1, 1]) = P_{\mathsf{S}_{n_d}\mathbf{X}_1^N | \mathbf{B}}(\mathbf{Y}_1^N \oplus \mathsf{S}_{n_c}\mathbf{X}_2^N | \mathbf{B} = [1, 1])$

(i4) $P_{\mathbf{Y}_2^N | \mathbf{X}_2^N, \mathbf{B}}(\mathbf{Y}_2^N | \mathbf{X}_2^N, \mathbf{B} = [1, 1]) = P_{\mathsf{S}_{n_c}\mathbf{X}_1^N | \mathbf{B}}(\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N | \mathbf{B} = [1, 1])$

(i5) $P_{\mathbf{Y}_1^N | \mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{B}}(\mathbf{Y}_1^N | \mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{B} = [1, 1]) = 1$

(i6)　$P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{X}_2^N,\mathbf{B}}(\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{X}_2^N,\mathbf{B}=[1,1])=1.$

Using (A115) and the identities (i1)–(i6), we obtain for (A100)

$$
\begin{aligned}
\epsilon_{11} = \Pr\Bigg\{ & -\frac{1}{N}\log P_{S_{n_d}\mathbf{X}_1^N|\mathbf{B}}(S_{n_d}\mathbf{X}_1^N|\mathbf{B}) + \frac{1}{N}\log P_{S_{n_c}\mathbf{X}_1^N|\mathbf{B}}(S_{n_c}\mathbf{X}_1^N|\mathbf{B}) \\
& -\frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) + \frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}) \le \Gamma \,\Big|\, \mathbf{B}=[1,1] \Bigg\}.
\end{aligned}
\tag{A116}
$$

Using (A108) in (A116), canceling the term $\log P_{S_{n_c}\mathbf{X}_1^N|\mathbf{B}}(S_{n_c}\mathbf{X}_1^N|\mathbf{B})$, and using that $\log P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}) \le 0$, we obtain the lower bound

$$
\epsilon_{11} \ge \Pr\Bigg\{ -\frac{1}{N}\log P_{\tilde{L}_{(n_d-n_c)^+}\mathbf{X}_1^N|S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{L}_{(n_d-n_c)^+}\mathbf{X}_1^N|S_{n_c}\mathbf{X}_1^N,\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \le \Gamma \,\Big|\, \mathbf{B}=[1,1] \Bigg\}. \tag{A117}
$$

The RHS of (A117) coincides with (A109) conditioned on $\mathbf{B}=[1,1]$. The proof then follows the one in Appendix C.2.2, with the probabilities and sup-entropy rates conditioned on $\mathbf{B}=[1,1]$ instead of $\mathbf{B}=[0,1]$.

Appendix C.3.3. Proof of Constraint (A8)

We begin this proof by using (A96) to write

$$
\begin{aligned}
i_1 + i_2 = & \log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b})} + \log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b})} \\
\overset{(a)}{=} & \log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,S_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|S_{n_c}\mathbf{x}_1^N,\mathbf{b})} + \log \frac{P_{\mathbf{X}_1^N|S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|S_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b})} \\
& + \log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,S_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|S_{n_c}\mathbf{x}_2^N,\mathbf{b})} + \log \frac{P_{\mathbf{X}_2^N|S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|S_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b})} \\
& - \log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,S_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{b})} - \log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,S_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,\mathbf{b})} \\
\overset{(b)}{=} & \log P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_1^N,S_{n_c}\mathbf{x}_1^N,\mathbf{b}) - \log P_{\mathbf{Y}_1^N|S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|S_{n_c}\mathbf{x}_1^N,\mathbf{b}) \\
& + \log P_{S_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}}(S_{n_c}\mathbf{x}_1^N|\mathbf{x}_1^N,\mathbf{b}) - \log P_{S_{n_c}\mathbf{X}_1^N|\mathbf{B}}(S_{n_c}\mathbf{x}_1^N|\mathbf{b}) \\
& + \log P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathbf{x}_2^N,S_{n_c}\mathbf{x}_2^N,\mathbf{b}) - \log P_{\mathbf{Y}_2^N|S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|S_{n_c}\mathbf{x}_2^N,\mathbf{b}) \\
& + \log P_{S_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}}(S_{n_c}\mathbf{x}_2^N|\mathbf{x}_2^N,\mathbf{b}) - \log P_{S_{n_c}\mathbf{X}_2^N|\mathbf{B}}(S_{n_c}\mathbf{x}_2^N|\mathbf{b}) \\
& - \log \frac{P_{S_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{Y}_1^N,\mathbf{B}}(S_{n_c}\mathbf{x}_1^N|\mathbf{x}_1^N,\mathbf{y}_1^N,\mathbf{b})}{P_{S_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(S_{n_c}\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{b})} - \log \frac{P_{S_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{Y}_2^N,\mathbf{B}}(S_{n_c}\mathbf{x}_2^N|\mathbf{x}_2^N,\mathbf{y}_2^N,\mathbf{b})}{P_{S_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(S_{n_c}\mathbf{x}_2^N|\mathbf{y}_2^N,\mathbf{b})}
\end{aligned}
\tag{A118}
$$

where (a) follows by adding and subtracting

$$
\frac{1}{N}\log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,S_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|S_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|S_{n_c}\mathbf{x}_1^N,\mathbf{b})} \quad\text{and}\quad \frac{1}{N}\log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,S_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|S_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|S_{n_c}\mathbf{x}_2^N,\mathbf{b})}
$$

and by rearranging terms. Step (b) follows by applying the Bayes rule and by decomposing the logarithm terms.

We analyze the second and the seventh terms in (A118). To this end, we define $n_- \triangleq \min\{(n_d-n_c)^+, n_c\}$ and $n_+ \triangleq \max\{(n_d-n_c)^+, n_c\}$ and apply the chain rule of probability to obtain

$$P_{\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})$$

$$= P_{\mathsf{S}_{n_-}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathsf{S}_{n_-}\mathbf{y}_1^N,\mathbf{b}) \quad \text{(A119)}$$

$$= P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathsf{S}_{n_-}\mathbf{y}_1^N,\mathbf{b})$$

and

$$P_{\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})$$

$$= P_{\mathsf{S}_{n_-}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathsf{S}_{n_-}\mathbf{y}_2^N,\mathbf{b}) \quad \text{(A120)}$$

$$= P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathsf{S}_{n_-}\mathbf{y}_2^N,\mathbf{b}).$$

The probabilities (A119) and (A120) were simplified by recalling the underlying distribution (A113). Indeed, we have w.p. 1 that $P_{\mathsf{S}_{n_-}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}) = 1$ and $P_{\mathsf{S}_{n_-}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}) = 1$, since $\mathsf{S}_{n_-}\mathbf{Y}_i^N$, $i = 1,2$ is not affected by interference, so it is determined by $\mathsf{S}_{n_c}\mathbf{X}_i^N$, $i = 1,2$. Similarly, we have that $P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}) = 1$ and $P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}) = 1$.

We next note that, for the underlying distribution in (A113), the following identities hold w.p. 1:

(i1) $\mathbf{Y}_1^N \oplus \mathsf{S}_{n_d}\mathbf{X}_1^N = \mathsf{S}_{n_c}\mathbf{X}_2^N$

(i2) $\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N = \mathsf{S}_{n_c}\mathbf{X}_1^N$

(i3) $P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B} = [1,1]) = 1$

(i4) $P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B} = [1,1]) = 1$

(i5) $P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B} = [1,1]) = P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathbf{Y}_1^N \oplus \mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B} = [1,1])$

(i6) $P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B} = [1,1]) = P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N|\mathbf{B} = [1,1])$

We combine these identities with (A100) and (A118)–(A120) to obtain

$$\epsilon_{11} = \Pr\left\{ \frac{1}{N}(\mathsf{i}_1 + \mathsf{i}_2) \leq \Gamma \,\middle|\, \mathbf{B} = [1,1] \right\}$$

$$= \Pr\left\{ \frac{1}{N}\left(\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}) - \log P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}) \right.\right.$$

$$- \log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}) + \log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B})$$

$$+ \log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}) - \log P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B})$$

$$\left.\left. - \log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}) + \log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}) \right) \leq \Gamma \,\middle|\, \mathbf{B} = [1,1] \right\} \quad \text{(A121)}$$

$$\geq \Pr\left\{ -\frac{1}{N}\log P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}) \right.$$

$$\left. -\frac{1}{N}\log P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}) \leq \Gamma \,\middle|\, \mathbf{B} = [1,1] \right\}$$

where in the last step we canceled the terms $\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B})$ and $\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B})$ and we used that, w.p. 1, $\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}) \leq 0$ and $\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}) \leq 0$.

By (A94) and (A95) in Lemma A4, the conditional sup-entropy rates satisfy

$$\overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_i^N|\mathsf{S}_{n_c}\mathbf{X}_i^N,\mathsf{S}_{n_-}\mathbf{Y}_i^N,\mathbf{B}) \leq \overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_i^N|\mathbf{B}) < \max\{(n_d - n_c)^+, n_c\}, \quad i = 1,2. \quad \text{(A122)}$$

Then, setting $\Gamma = 2\max\{(n_d - n_c)^+, n_c\} + 2\delta$ for some arbitrary $\delta > 0$, we obtain from (A99) that (A121) can be lower-bounded by

$$
\begin{aligned}
\epsilon_{11} \geq \Pr\Bigg\{ &-\frac{1}{N}\log P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}) \\
&< \overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}) + \delta \,\Big|\, \mathbf{B} = [1,1]\Bigg\} \\
-\Pr\Bigg\{ &\frac{1}{N}\log P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_-}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}) \\
&\geq \overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}) + \delta \,\Big|\, \mathbf{B} = [1,1]\Bigg\}.
\end{aligned}
\tag{A123}
$$

By the definition of $\overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B})$, the fist probability on the RHS of (A123) tends to 1 as $N \to \infty$. Similarly, by the definition of $\overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B})$, the second probability on the RHS of (A123) tends to 0 as $N \to \infty$. This demonstrates that if $\Gamma > 2\max\{(n_d - n_c)^+, n_c\}$, then the lower bound in (A121) tends to 1 as $N \to \infty$. Thus, $\epsilon_{11} \to 0$ as $N \to \infty$ only if

$$\Gamma \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{A124}$$

Appendix D. Achievability for Local CSIRT

In this appendix we present the achievability schemes for local CSIRT.

Appendix D.1. Very Weak Interference

The sum rate (29) coincides with that of local CSIR, which in this interference region is equal to the sum rate of global CSIRT. The achievability scheme presented in Appendix B.2.1 is thus optimal for local CSIRT and VWI.

Appendix D.2. Weak Interference

We follow a random-coding argument where the codebooks of Tx_1 and Tx_2 are drawn i.i.d. at random according to the distribution depicted in Figure A7. Specifically, we divide the transmitted signal by Tx_1 into three regions. For each symbol (corresponding to a coherence block) we denote the bits in regions $\boxed{A}$, $\boxed{B}$ and $\boxed{C}$ by X_1^A, X_1^B and X_1^C, respectively. In each region the bits are i.i.d. but they follow a different distribution.

- Regions $\boxed{A}$ and $\boxed{C}$: The bits X_1^A and X_1^C are i.i.d. with marginal probability mass function (pmf)

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A125}$$

- Region $\boxed{B}$: The bits X_1^B are i.i.d. with marginal pmf

$$
\begin{aligned}
P_{X_1|B_1}(1|0) &= p_1 & \text{(A126)} \\
P_{X_1|B_1}(1|1) &= p_2 & \text{(A127)} \\
P_{X_1}(1) &= p_3 = (1-p)p_1 + pp_2. & \text{(A128)}
\end{aligned}
$$

We further assume that X_1^A, X_1^B and X_1^C are mutually independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions but with probabilities q_i instead of p_i, with $i = 1, 2$. Evaluating $I(X_1; Y_1|B_1)$ for these distributions, it follows that user 1 achieves the rate

$$
\begin{aligned}
R_1 =&(1-p)\left[(n_d-n_c)H_b(\tfrac{1}{2})+(2n_c-n_d)H_b(p_1)+(n_d-n_c)\right]+p(n_d-n_c)H_b(\tfrac{1}{2})\\
&+p(2n_c-nd)\left[H_{\text{sum}}(p_2,\tfrac{1}{2})-H_b(q_3)\right]+p(2n_d-3n_c)(H_{\text{sum}}(\tfrac{1}{2},\tfrac{1}{2})-H_b(\tfrac{1}{2}))\\
&+p(2n_c-n_d)(H_{\text{sum}}(\tfrac{1}{2},q_3)-H_b(q_3))\\
=&(n_d-n_c)+(1-p)\left[(n_d-n_c)+(2n_c-n_d)H_b(p_1)\right]+p(2n_c-n_d)(1-H_b(q_3)).
\end{aligned}
\tag{A129}
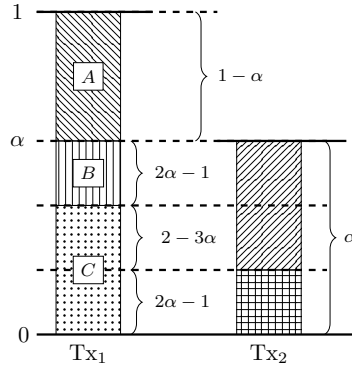$$

Similarly, for user 2, we obtain (31).

Figure A7. Normalized signal levels at Rx_1 (WI).

Appendix D.3. Moderate Interference

We follow along similar lines to obtain the achievable rates for MI. However, in contrast to WI, for MI we need to consider different input distributions, depending on the value of α. In the proofs, we shall make use of the following auxiliary results, which can be proven by direct evaluation of the entropies considered.

Lemma A5. *Let X and $\tilde{X}$ be two binary random variables with joint pmf $P_{X\tilde{X}}(0,0)=P_{X\tilde{X}}(1,1)=\frac{\eta}{2}$, and $P_{X\tilde{X}}(0,1)=P_{X\tilde{X}}(1,0)=\frac{1-\eta}{2}$. Then,*

$$
H(X|\tilde{X})=H(\tilde{X}|X)=H_b(\eta).
\tag{A130}
$$

Lemma A6. *Let $X,\tilde{X}$ and B be binary random variables with joint pmf $P_{X\tilde{X}B}(0,0,0)=P_{X\tilde{X}B}(1,1,0)=\frac{\eta_1}{2}(1-p)$, $P_{X\tilde{X}B}(0,1,0)=P_{X\tilde{X}B}(1,0,0)=\frac{1-\eta_1}{2}(1-p)$, $P_{X\tilde{X}B}(0,0,1)=P_{X\tilde{X}B}(1,1,1)=\frac{\eta_2}{2}p$, and $P_{X\tilde{X}B}(0,1,1)=P_{X\tilde{X}B}(1,0,1)=\frac{1-\eta_2}{2}p$. Then,*

$$
H(\tilde{X}|X,B)=(1-p)H_b(\eta_1)+pH_b(\eta_2)
\tag{A131}
$$

and

$$
H(\tilde{X}|X)=H_b\left((1-p)\eta_1+p\eta_2\right).
\tag{A132}
$$

Lemma A7. *Let X_1 and $\tilde{X}_1$ be two binary random variables with joint pmf $P_{X_1\tilde{X}_1}(0,0)=P_{X_1\tilde{X}_1}(1,1)=\frac{\eta_1}{2}$ and $P_{X_1\tilde{X}_1}(0,1)=P_{X_1\tilde{X}_1}(1,0)=\frac{1-\eta_1}{2}$. Similarly, let the pair of binary random variables X_2 and $\tilde{X}_2$ be independent of X_1 and $\tilde{X}_1$ have the same joint pmf but with parameter η_2. Further let $Z\sim Ber(p_z)$. Then,*

$$
H(X_1|\tilde{X}_1\oplus\tilde{X}_2,X_2)=H(\tilde{X}_1\oplus\tilde{X}_2|X_1,X_2)=H_{sum}(\eta_1,\eta_2)
\tag{A133}
$$

and

$$
H(X_1\oplus Z|\tilde{X}_1\oplus\tilde{X}_2,X_2)=H_{sum}(p_z,\eta_1(1-\eta_2)+\eta_2(1-\eta_1)).
\tag{A134}
$$

To derive the achievable rates for MI, we again follow a random-coding argument where the codebooks are drawn i.i.d. at random. We next describe the input distributions for different values of α:

Appendix D.3.1. MI, $\frac{2}{3} < \alpha \le \frac{3}{4}$

Consider the regions shown in Figure A8 for the received signal at Rx$_1$. For the transmitted signal X_1, we denote the bits in region $\boxed{j}$ by X_1^j, $j = \{A, \dots, F\}$. In each of these regions we consider the following input distributions:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: We group the bits X_1^A and $X_1^{\tilde{A}}$ in pairs, and we let each of these pairs $(X_1, \tilde{X}_1)$ be i.i.d. and have the distribution from Lemma A6 with $\eta_2 = 1$, i.e., their marginal pmf is

$$P_{X_1 \tilde{X}_1 | B_1}(0,0|0) = P_{X_1 \tilde{X}_1 | B_1}(1,1|0) \quad = \quad \frac{\eta_1}{2} \tag{A135}$$

$$P_{X_1 \tilde{X}_1 | B_1}(0,1|0) = P_{X_1 \tilde{X}_1 | B_1}(1,0|0) \quad = \quad \frac{1 - \eta_1}{2} \tag{A136}$$

$$P_{X_1 \tilde{X}_1 | B_1}(0,0|1) = P_{X_1 \tilde{X}_1 | B_1}(1,1|1) \quad = \quad \frac{1}{2} \tag{A137}$$

$$P_{X_1 \tilde{X}_1 | B_1}(0,1|1) = P_{X_1 \tilde{X}_1 | B_1}(1,0|1) \quad = \quad 0 \tag{A138}$$

$$P_{\tilde{X}_1 | X_1}(1|1) \quad = \quad \tilde{\eta} = p + \eta_1(1 - p). \tag{A139}$$

  where $\frac{1}{2} \le \eta_1 \le 1$.
- Regions $\boxed{B}$ and $\boxed{F}$: The bits X_1^B and X_1^F are i.i.d. with marginal pmf

$$P_{X_1 | B_1}(1|0) = P_{X_1 | B_1}(1|1) \quad = \quad \tfrac{1}{2}. \tag{A140}$$

- Region $\boxed{C}$: The bits X_1^C are i.i.d. with marginal pmf

$$P_{X_1 | B_1}(1|0) \quad = \quad p_1 \tag{A141}$$

$$P_{X_1 | B_1}(1|1) \quad = \quad p_2 \tag{A142}$$

$$P_{X_1}(1) \quad = \quad p_3 = (1 - p)p_1 + p p_2. \tag{A143}$$

- Region $\boxed{D}$: The bits X_1^D are i.i.d. with marginal pmf

$$P_{X_1 | B_1}(1|0) \quad = \quad \tilde{p}_1 \tag{A144}$$

$$P_{X_1 | B_1}(1|1) \quad = \quad \tilde{p}_2 \tag{A145}$$

$$P_{X_1}(1) \quad = \quad \tilde{p}_3 = (1 - p)\tilde{p}_1 + p\tilde{p}_2. \tag{A146}$$

- Region $\boxed{E}$: The bits X_1^E are i.i.d. with marginal pmf

$$P_{X_1 | B_1}(1|0) \quad = \quad \hat{p}_1 \tag{A147}$$

$$P_{X_1 | B_1}(1|1) \quad = \quad 0 \tag{A148}$$

$$P_{X_1}(1) \quad = \quad \hat{p}_3 = (1 - p)\hat{p}_1. \tag{A149}$$

Furthermore, we assume that X_1^j, $j = \{A, \dots, F\}$ are independent. For user 2, the input distributions coincide with that of user 1 in the corresponding regions, but with parameters q_i instead of p_i, $\tilde{q}_i$ instead of $\tilde{p}_i$, $\hat{q}_1$ instead of $\hat{p}_1$, and γ_i instead of η_i.
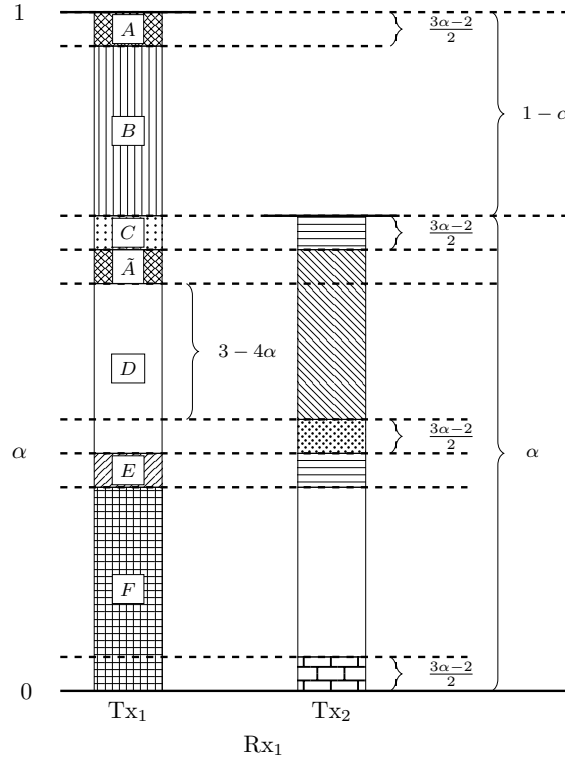
Figure A8. Normalized signal levels at Rx_1 (MI) for $\frac{2}{3} < \alpha \leq \frac{3}{4}$.

From the random-coding argument, we know that the rate $R_1 = \frac{1}{N} I(\boldsymbol{X}_1^K; \boldsymbol{Y}_1^K | B_1)$ is achievable. Since the distributions considered are temporally i.i.d., it suffices to evaluate $I(\boldsymbol{X}_1; \boldsymbol{Y}_1 | B_1)$ for one coherence block, obtaining

$$
\begin{aligned}
TR_1 =\, & I(\boldsymbol{X}_1; \boldsymbol{Y}_1^A | B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^{\tilde{A}} | \boldsymbol{Y}_1^A, B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^E | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^C | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, B_1) \\
& + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^B | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, \boldsymbol{Y}_1^C, B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^D | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, \boldsymbol{Y}_1^C, \boldsymbol{Y}_1^B, B_1) \\
& + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^F | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, \boldsymbol{Y}_1^C, \boldsymbol{Y}_1^B, \boldsymbol{Y}_1^D, B_1) \\
=\, & (1 - p)[H(\boldsymbol{X}_1^A | B_1 = 0) + H(\boldsymbol{X}_1^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 0) + H(\boldsymbol{X}_1^E | B_1 = 0) + H(\boldsymbol{X}_1^C | B_1 = 0) \\
& + H(\boldsymbol{X}_1^B | B_1 = 0) + H(\boldsymbol{X}_1^D | B_1 = 0) + H(\boldsymbol{X}_1^F | B_1 = 0)] \\
& + p[H(\boldsymbol{X}_1^A | B_1 = 1) + H(\boldsymbol{X}_1^{\tilde{A}} \oplus \boldsymbol{X}_2^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 1) - H(\boldsymbol{X}_2^{\tilde{A}}) + H(\boldsymbol{X}_1^E \oplus \boldsymbol{X}_2^E | B_1 = 1) \\
& - H(\boldsymbol{X}_2^E) + H(\boldsymbol{X}_1^C \oplus \boldsymbol{X}_2^C | \boldsymbol{X}_1^E \oplus \boldsymbol{X}_2^E, B_1 = 1) - H(\boldsymbol{X}_2^C | \boldsymbol{X}_2^E) + H(\boldsymbol{X}_1^B | B_1 = 1) \\
& + H(\boldsymbol{X}_1^D \oplus \boldsymbol{X}_2^D | B_1 = 1) - H(\boldsymbol{X}_2^D) + H(\boldsymbol{X}_1^F \oplus \boldsymbol{X}_2^F | B_1 = 1) - H(\boldsymbol{X}_2^F)].
\end{aligned}
\tag{A150}
$$

By Lemma A6, we have that

$$
H(\boldsymbol{X}_1^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 0) = T \frac{3n_c - 2n_d}{2} H_b(\eta_1).
\tag{A151}
$$

Furthermore, by Lemma A7, we have that

$$
\begin{aligned}
H(\boldsymbol{X}_1^C \oplus \boldsymbol{X}_2^C | \boldsymbol{X}_1^E \oplus \boldsymbol{X}_2^E, B_1 = 1) &= H(\boldsymbol{X}_1^C \oplus \boldsymbol{X}_2^C | \boldsymbol{X}_2^E, B_1 = 1) \\
&= T \frac{3n_c - 2n_d}{2} H_{\text{sum}}(p_2, \tilde{\gamma})
\end{aligned}
\tag{A152}
$$

because for the bits $\boldsymbol{X}_1^C$, $P_{X_1^C | B_1}(1|1) = 0$. Similarly, we have

$$
H_{\text{sum}}(\boldsymbol{X}_1^{\tilde{A}} \oplus \boldsymbol{X}_2^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 1) = T \frac{3n_c - 2n_d}{2} H_{\text{sum}}(1, \tfrac{1}{2}) = T \frac{3n_c - 2n_d}{2}.
\tag{A153}
$$

The terms in the other regions follow analogously. Therefore, using (A150) we obtain the rate

$$
\begin{aligned}
R_1 =& (n_d - n_c) + (1 - p)\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_b(\eta_1) + H_b(\hat{p}_1) + H_b(p_1)) + \left(\tfrac{4n_d - 5n_c}{2}\right)H_b(\tilde{p}_1) + (n_d - n_c)\right] \\
&+ p\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(1 + H_{\text{sum}}(p_2, \tilde{\gamma}) - H_b(\tilde{\gamma}) + H_{\text{sum}}(\tilde{p}_2, q_3) - H_b(q_3) - H_b(\hat{q}_3)) \\
&+ \left(\tfrac{4n_d - 5n_c}{2}\right)(1 - H_b(\tilde{q}_3))\right].
\end{aligned}
\tag{A154}
$$

Similarly, user 2 achieves the rate (33).

Appendix D.3.2. MI, $\frac{3}{4} \le \alpha \le \frac{4}{5}$

We use a similar transmission strategy as for the case where $\frac{2}{3} \le \alpha \le \frac{3}{4}$ (Section D.3.1), but where the regions have different sizes; see Figure A9a. Following the same steps as in Section D.3.1, we obtain the achievable rates (34) for R_1 and (35) for R_2.
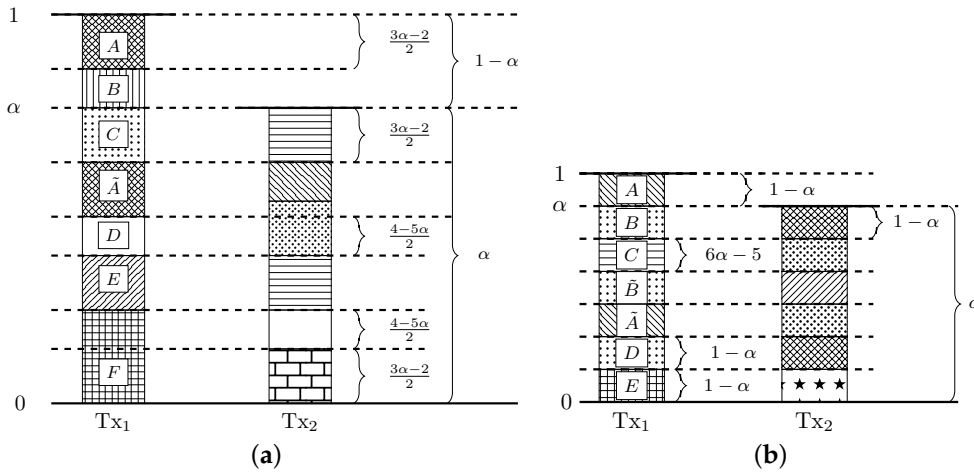

Figure A9. Normalized signal levels at Rx$_1$. (**a**) (MI) for $\frac{3}{4} \le \alpha \le \frac{4}{5}$; (**b**) (MI) for $\alpha = \frac{6}{7}$.

Appendix D.3.3. MI, $\alpha = \frac{6}{7}$

In this subsection we consider the particular case $\alpha = \frac{6}{7}$. The proposed achievability scheme features two nested regions with a certain correlation. In particular, we consider the division of the bit-pipes for the transmitted signal Tx$_1$ in the subregions shown in Figure A9b. The input distributions considered in each of these regions are described next (for Tx$_2$, we shall consider the same input distributions parametrized by q_i, $\hat{q}_1$, γ_1 and γ', instead of p_i, $\hat{p}_1$, η_1 and η'):

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits X_1^A and $X_1^{\tilde{A}}$ are grouped in i.i.d. pairs with the marginal pmf given by (A135)–(A139).
- Regions $\boxed{B}$ and $\boxed{\tilde{B}}$: The bits X_1^B and $X_1^{\tilde{B}}$ are grouped in i.i.d. pairs with marginal pmf

$$
P_{X_1 \tilde{X}_1 | B_1}(0,0|0) = P_{X_1 \tilde{X}_1 | B_1}(1,1|0) = \frac{\eta'}{2}
\tag{A155}
$$

$$
P_{X_1 \tilde{X}_1 | B_1}(0,1|0) = P_{X_1 \tilde{X}_1 | B_1}(1,0|0) = \frac{1 - \eta'}{2}
\tag{A156}
$$

$$
P_{X_1 \tilde{X}_1 | B_1}(0,0|1) = P_{X_1 \tilde{X}_1 | B_1}(1,1|1) = \frac{\eta'}{2}
\tag{A157}
$$

$$
P_{X_1 \tilde{X}_1 | B_1}(0,1|1) = P_{X_1 \tilde{X}_1 | B_1}(1,0|1) = \frac{1 - \eta'}{2}
\tag{A158}
$$

$$
P_{\tilde{X}_1 | X_1}(1|1) = \eta'
\tag{A159}
$$

where $\frac{1}{2} \leq \eta' \leq 1$.

- Region $\boxed{C}$: The bits X_1^C are i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) = p_1 \tag{A160}$$
$$P_{X_1|B_1}(1|1) = p_2 \tag{A161}$$
$$P_{X_1}(1) = p_3 = (1-p)p_1 + pp_2. \tag{A162}$$

- Region $\boxed{D}$: The bits X_1^D are i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) = \hat{p}_1 \tag{A163}$$
$$P_{X_1|B_1}(1|1) = 0 \tag{A164}$$
$$P_{X_1}(1) = \hat{p}_3 = (1-p)\hat{p}_1. \tag{A165}$$

- Region $\boxed{E}$: The bits X_1^E are i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A166}$$

Furthermore, we assume that X_{i}^{j}, i=1,2, $j = \{A, B, C, D, E\}$ are mutually independent. For the input distributions described above, we obtain for user 1 that

$$
\begin{aligned}
TR_1 =& I(X_1; Y_1^A | B_1) + I(X_1; Y_1^{\tilde{A}} | Y_1^A, B_1) + I(X_1; Y_1^D | Y_1^A, Y_1^{\tilde{A}}, B_1) + I(X_1; Y_1^B | Y_1^A, Y_1^{\tilde{A}}, Y_1^D, B_1) \\
&+ I(X_1; Y_1^{\tilde{B}} | Y_1^A, Y_1^{\tilde{A}}, Y_1^D, Y_1^B, B_1) + I(X_1; Y_1^C | Y_1^A, Y_1^{\tilde{A}}, Y_1^D, Y_1^B, Y_1^{\tilde{B}}, B_1) \\
&+ I(X_1; Y_1^E | Y_1^A, Y_1^{\tilde{A}}, Y_1^D, Y_1^B, Y_1^{\tilde{B}}, Y_1^C, B_1) \\
=& (1-p)[H(X_1^A | B_1 = 0) + H(X_1^{\tilde{A}} | X_1^A, B_1 = 0) + H(X_1^D | B_1 = 0) \\
&+ H(X_1^B | B_1 = 0) + H(X_1^{\tilde{B}} | X_1^B, B_1 = 0) + H(X_1^C | B_1 = 0) + H(X_1^E | B_1 = 0)] \\
&+ p[H(X_1^A | B_1 = 1) + H(X_1^{\tilde{A}} \oplus X_2^{\tilde{A}} | X_1^A, B_1 = 1) - H(X_2^{\tilde{A}}) + H(X_1^D \oplus X_2^D | B_1 = 1) \\
&- H(X_2^D) + H(X_1^B \oplus X_2^B | X_1^D \oplus X_2^D, B_1 = 1) - H(X_2^B | X_2^D) \\
&+ H(X_1^{\tilde{B}} \oplus X_2^{\tilde{B}} | X_1^B \oplus X_2^B, X_1^D \oplus X_2^D, B_1 = 1) - H(X_2^{\tilde{B}}) \\
&+ H(X_1^C \oplus X_2^C | X_1^{\tilde{A}} \oplus X_2^{\tilde{A}}, X_1^A, B_1 = 1) - H(X_2^C | X_2^{\tilde{A}}) + H(X_1^E \oplus X_2^E | B_1 = 1) \\
&- H(X_2^E)].
\end{aligned}
\tag{A167}
$$

We next evaluate the different terms in (A167) by applying Lemmas A6 and A7 to obtain

$$H(X_1^{\tilde{A}} | X_1^A, B_1 = 0) = T(n_d - n_c)H_b(\eta_1) \tag{A168}$$
$$H(X_1^{\tilde{B}} | X_1^B, B_1 = 0) = T(n_d - n_c)H_b(\eta') \tag{A169}$$
$$H(X_1^{\tilde{A}} \oplus X_2^{\tilde{A}} | X_1^A, B_1 = 1) = T(n_d - n_c)H_{\text{sum}}(1, \tfrac{1}{2}) = T(n_d - n_c). \tag{A170}$$

Similarly, using Lemma A7, and since for X_1^D we have that $P_{X_1|B_1}(0|1) = 1$, we obtain

$$
\begin{aligned}
H(X_1^{\tilde{B}} \oplus X_2^{\tilde{B}} | X_1^B \oplus X_2^B, X_1^D \oplus X_2^D, B_1 = 1) =& H(X_1^{\tilde{B}} \oplus X_2^{\tilde{B}} | X_1^B \oplus X_2^B, X_2^D, B_1 = 1) \\
=& T(n_d - n_c)H_{\text{sum}}(q_3, \eta'(1 - \tilde{\gamma}) + (1 - \eta')\tilde{\gamma}).
\end{aligned}
\tag{A171}
$$

Combining (A168)–(A171) with (A167) yields

$$
\begin{aligned}
R_1 =& (n_d - n_c) + (1-p)\big[(6n_c - 5n_d)H_b(p_1) + (n_d - n_c)\left(2 + H_b(\eta_1) + H_b(\eta') + H_b(\hat{p}_1)\right)\big] \\
&+ p\big[(n_d - n_c)\left(2 - H_b(\tilde{\gamma}) - H_b(\hat{q}_3) + H_{\text{sum}}(\eta'(1 - \tilde{\gamma}) + (1 - \eta')\tilde{\gamma}, q_3) - H_b(q_3)\right) \\
&+ (6n_c - 5n_d)\left(H_{\text{sum}}(p_2, \gamma') - H_b(\gamma')\right)\big].
\end{aligned}
\tag{A172}
$$

Following along similar lines, it can be shown that user 2 achieves the rate

$$
\begin{aligned}
R_2 =& (n_d - n_c) + (1 - p)\left[(6n_c - 5n_d)H_b(q_1) + (n_d - n_c)\left(2 + H_b(\gamma_1) + H_b(\gamma') + H_b(\hat{q}_1)\right)\right] \\
&+ p\left[(n_d - n_c)\left(2 - H_b(\tilde{\eta}) - H_b(\hat{p}_3) + H_{\mathrm{sum}}(\gamma'(1 - \tilde{\eta}) + (1 - \gamma')\tilde{\eta}, p_3) - H_b(p_3)\right)\right. \\
&\left. + (6n_c - 5n_d)\left(H_{\mathrm{sum}}(q_2, \eta') - H_b(\eta')\right)\right].
\end{aligned}
\tag{A173}
$$

Appendix D.3.4. MI, $\frac{4}{5} < \alpha < \frac{6}{7}$

We consider the input distribution depicted in Figure A10a with:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(X_1^A, X_1^{\tilde{A}})$ are i.i.d. with marginal pmf given by (A135)-(A139).
- Regions $\boxed{B}$ and $\boxed{\tilde{B}}$: The bits $(X_1^B, X_1^{\tilde{B}})$ are i.i.d. with marginal pmf given by (A155)-(A159).
- Region $\boxed{C}$: The bits X_1^C are i.i.d. with marginal pmf

$$
\begin{aligned}
P_{X_1|B_1}(1|0) &= p_1 && \text{(A174)} \\
P_{X_1|B_1}(1|1) &= p_2 && \text{(A175)} \\
P_{X_1}(1) &= p_3 = (1 - p)p_1 + pp_2. && \text{(A176)}
\end{aligned}
$$

- Region $\boxed{D}$: The bits X_1^D are i.i.d. with marginal pmf

$$
\begin{aligned}
P_{X_1|B_1}(1|0) &= \hat{p}_1 && \text{(A177)} \\
P_{X_1|B_1}(1|1) &= 0 && \text{(A178)} \\
P_{X_1}(1) &= \hat{p}_3 = (1 - p)\hat{p}_1. && \text{(A179)}
\end{aligned}
$$

- Region $\boxed{E}$: The bits X_1^E are i.i.d. with marginal pmf

$$
P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}
\tag{A180}
$$

Furthermore, we assume that X_1^j, $j = \{A, B, C, D, E\}$ are independent. For Tx$_2$, the input distributions coincide with that of Tx$_1$ in the corresponding regions, but with parameters q_i instead of p_i, $\hat{q}_1$ instead of $\hat{p}_1$, γ_i instead of η_i and γ' instead of η'. Following similar steps as in the previous sections, we obtain (36) for R_1 and (37) for R_2.

Appendix D.3.5. MI, $\frac{6}{7} < \alpha < 1$

The transmission strategy is similar to the one for $\frac{4}{5} < \alpha < \frac{6}{7}$ (Section D.3.4), but with different sizes for the regions $\boxed{A}$ - $\boxed{E}$, see Figure A10b. Following similar steps as in previous sections, we obtain (38) for R_1 and (39) for R_2.

Appendix D.4. Strong Interference

To obtain the achievable rates for SI, we again need to consider different input distributions, depending on the value of α.

Appendix D.4.1. SI, $1 \leq \alpha \leq \frac{6}{5}$

We consider the input distribution depicted in Figure A11a with:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(X_1^A, X_1^{\tilde{A}})$ are i.i.d. with marginal pmf given by (A135)–(A139).
- Regions $\boxed{B}$ and $\boxed{\tilde{B}}$: The bits $(X_1^B, X_1^{\tilde{B}})$ are i.i.d. with marginal pmf given by (A155)–(A159).

- Region $\boxed{C}$: The bits X_1^C are i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) \;=\; p_1 \tag{A181}$$
$$P_{X_1|B_1}(1|1) \;=\; p_2 \tag{A182}$$
$$P_{X_1}(1) \;=\; p_3 = (1-p)p_1 + pp_2. \tag{A183}$$

Furthermore, we assume that $X_1^j, j = \{A, B, C\}$ are independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions, but with parameters q_i instead of p_i, γ_1 instead of η_1 and γ' instead of η'. Following similar steps as in previous sections, we obtain the achievable rate pair (40) and (41).
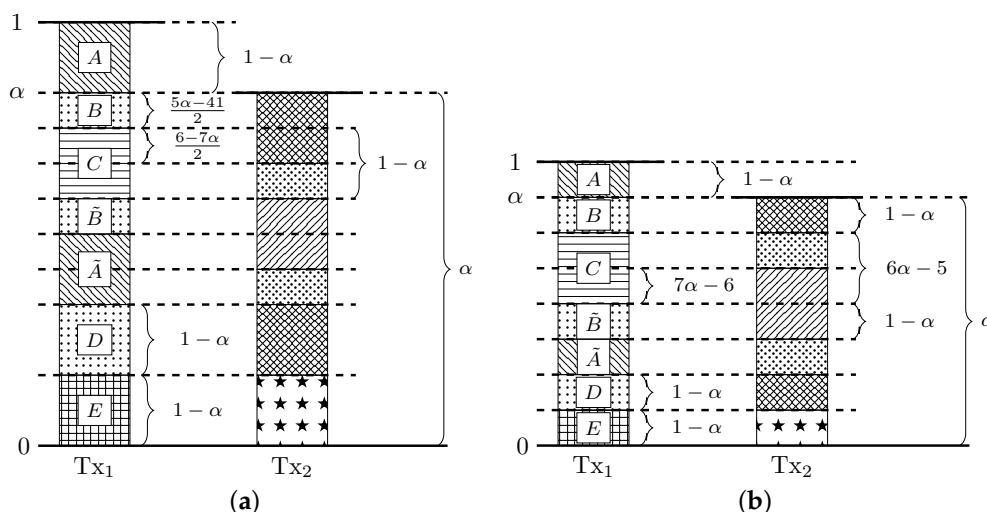

Figure A10. Normalized signal levels at Rx_1. (**a**) (MI) for $\frac{4}{5} \le \alpha \le \frac{6}{7}$; (**b**) (MI) for $\frac{6}{7} \le \alpha \le 1$.
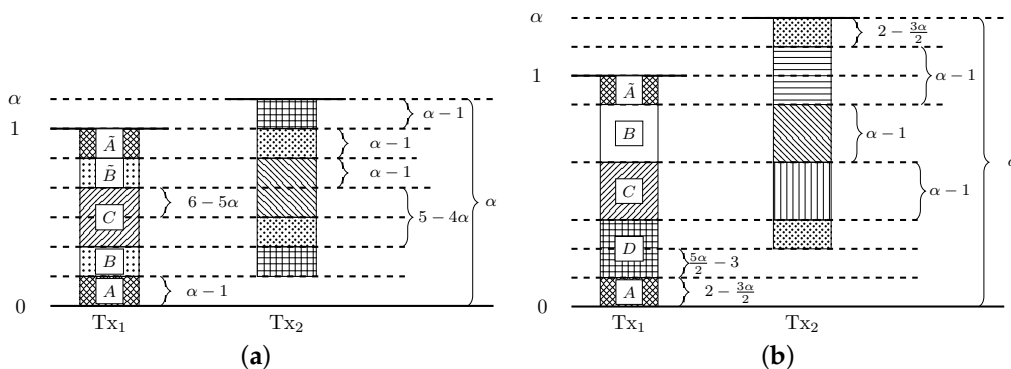

Figure A11. Normalized signal levels at Rx_1. (**a**) (SI) for $1 \le \alpha \le \frac{6}{5}$.; (**b**) (SI) for $\frac{6}{5} \le \alpha \le \frac{4}{3}$.

Appendix D.4.2. SI, $\frac{6}{5} \le \alpha \le \frac{4}{3}$

We consider the input distribution depicted in Figure A11b with the following distributions:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(X_1^A, X_1^{\tilde{A}})$ are i.i.d. with marginal pmf given by (A135)–(A139).
- Regions $\boxed{B}$ and $\boxed{D}$: The bits X_1^B and X_1^D are independent and temporally i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A184}$$

- Region $\boxed{C}$: The bits $\boldsymbol{X}_1^C$ are i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) \quad = \quad p_1 \tag{A185}$$
$$P_{X_1|B_1}(1|1) \quad = \quad p_2 \tag{A186}$$
$$P_{X_1}(1) \quad = \quad p_3 = (1-p)p_1 + pp_2. \tag{A187}$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B, C, D\}$ are independent. For Tx$_2$, the input distributions coincide with that of Tx$_1$ in the corresponding regions, but with parameters q_i instead of p_i, $\hat{q}_1$ instead of $\hat{p}_1$ and γ_1 instead of η_1. Following similar steps as in previous sections, we obtain the achievable rate pair (42) and (43).

Appendix D.4.3. SI, $\frac{4}{3} \le \alpha \le \frac{3}{2}$

We consider the input distribution depicted in Figure A12a with the following distributions:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(\boldsymbol{X}_1^A, \boldsymbol{X}_1^{\tilde{A}})$ are i.i.d. with marginal pmf given by (A135)–(A139).
- Regions $\boxed{B}$, $\boxed{C}$, $\boxed{E}$ and $\boxed{F}$: The bits $\boldsymbol{X}_1^B$, $\boldsymbol{X}_1^C$, $\boldsymbol{X}_1^E$ and $\boldsymbol{X}_1^F$ are independent and temporally i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A188}$$

- Region $\boxed{D}$: The bits $\boldsymbol{X}_1^D$ are i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) \quad = \quad \hat{p}_1 \tag{A189}$$
$$P_{X_1|B_1}(1|1) \quad = \quad \hat{p}_1 \tag{A190}$$
$$P_{X_1}(1) \quad = \quad \hat{p}_3 = \hat{p}_1. \tag{A191}$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B, C, D, E, F\}$ are independent. For Tx$_2$, the input distributions coincide with that of Tx$_1$ in the corresponding regions, but with parameters q_i instead of p_i and γ_1 instead of η_1. Following similar steps as in previous sections, we obtain an achievable rate pair for $\frac{4}{3} < \alpha \le \frac{3}{2}$ which is given by (44) and (45).

Appendix D.4.4. SI, $\frac{3}{2} \le \alpha \le 2$

We consider the input distribution depicted in Figure A12b with the following distributions:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(\boldsymbol{X}_1^A, \boldsymbol{X}_1^{\tilde{A}})$ are i.i.d. with marginal pmf given by (A135)–(A139).
- Region $\boxed{B}$: The bits are i.i.d. with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A192}$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B\}$ are independent. For Tx$_2$, the input distributions coincide with that of Tx$_1$ in the corresponding regions, but with parameters q_i instead of p_i, $\hat{q}_1$ instead of $\hat{p}_1$ and γ_1 instead of η_1. Proceeding as in the previous sections we obtain the achievable rate pair (46) and (47).

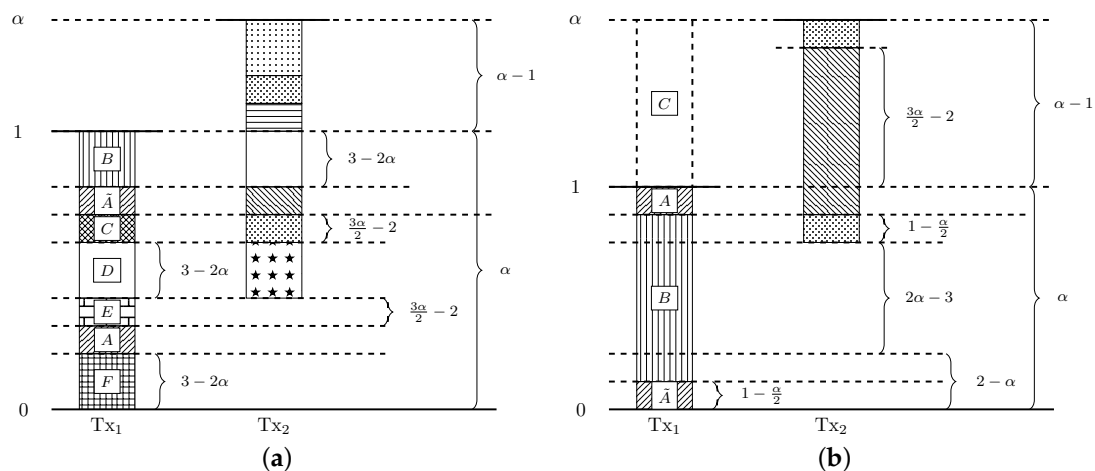**Figure A12.** Normalized signal levels at Rx_1. (**a**) (SI) for $\frac{4}{3} \leq \alpha \leq \frac{3}{2}$; (**b**) (SI) for $\frac{3}{2} \leq \alpha \leq 2$.

References

1. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
2. Khude, N.; Prabhakaran, V.; Viswanath, P. Harnessing bursty interference. In Proceedings of the 2009 IEEE Information Theory Workshop on Networking and Information Theory (ITWNIT), Volos, Greece, 12–10 June 2009; pp. 13–16, doi:10.1109/ITWNIT.2009.5158532. [CrossRef]
3. Khude, N.; Prabhakaran, V.; Viswanath, P. Opportunistic interference management. In Proceedings of the 2009 IEEE International Symposium on Information Theory (ISIT), Seoul, South Korea, 28 June–3 July 2009; pp. 2076–2080, doi:10.1109/ISIT.2009.5205335. [CrossRef]
4. Korner, J.; Marton, K. General broadcast channels with degraded message sets. *IEEE Trans. Inf. Theory* **1977**, *23*, 60–64, doi:10.1109/TIT.1977.1055655. [CrossRef]
5. Diggavi, S.N.; Tse, D.N.C. On opportunistic codes and broadcast codes with degraded message sets. In Proceedings of the 2006 IEEE Information Theory Workshop (ITW), Punta del Este, Uruguay, 13–17 March 2006; pp. 227–231, doi:10.1109/ITW.2006.1633817. [CrossRef]
6. Yi, X.; Sun, H. Opportunistic treating interference as noise. *arXiv* **2018**, arXiv:1808.08926.
7. Wang, I.H.; Suh, C.; Diggavi, S.; Viswanath, P. Bursty interference channel with feedback. In Proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, 7–12 July 2013; pp. 21–25, doi:10.1109/ISIT.2013.6620180. [CrossRef]
8. Mishra, S.; Wang, I.H.; Diggavi, S.N. Harnessing bursty interference in multicarrier systems with output feedback. *IEEE Trans. Inf. Theory* **2017**, *63*, 4430–4452. doi:10.1109/TIT.2017.2697866. [CrossRef]
9. Mishra, S.; Wang, I.H.; Diggavi, S.N. Opportunistic interference management for multicarrier systems. In Proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, 7–12 July 2013; pp. 389–393, doi:10.1109/ISIT.2013.6620254. [CrossRef]
10. Vahid, A.; Maddah-Ali, M.A.; Avestimehr, A.S. Interference channel with binary fading: Effect of delayed network state information. In Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 28–30 September 2011; pp. 894–901, doi:10.1109/Allerton.2011.6120261. [CrossRef]
11. Vahid, A.; Maddah-Ali, M.A.; Avestimehr, A.S. Binary fading interference channel with no CSIT. In Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, 29 June–4 July 2014; pp. 666–670, doi:10.1109/ISIT.2014.6874916. [CrossRef]
12. Vahid, A.; Maddah-Ali, M.A.; Avestimehr, A.S. Capacity results for binary fading interference channels with delayed CSIT. *IEEE Trans. Inf. Theory* **2014**, *60*, 6093–6130, doi:10.1109/TIT.2014.2345371. [CrossRef]
13. Vahid, A.; Calderbank, R. When does spatial correlation add value to delayed channel state information? In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 2624–2628, doi:10.1109/ISIT.2016.7541774. [CrossRef]

14. Vahid, A.; Maddah-Ali, M.A.; Avestimehr, A.S.; Zhu, Y. Binary fading interference channel with no CSIT. *IEEE Trans. Inf. Theory* **2017**, *63*, 3565–3578, doi:10.1109/TIT.2017.2688335. [CrossRef]

15. Avestimehr, A.S.; Diggavi, S.N.; Tse, D.N.C. Wireless network information flow: A deterministic approach. *IEEE Trans. Inf. Theory* **2011**, *57*, 1872–1905, doi:10.1109/TIT.2011.2110110. [CrossRef]

16. El Gamal, A.; Costa, M. The capacity region of a class of deterministic interference channels. *IEEE Trans. Inf. Theory* **1982**, *28*, 343–346, doi:10.1109/TIT.1982.1056466. [CrossRef]

17. Bresler, G.; Tse, D. The two-user Gaussian interference channel: A deterministic view. *Trans. Emerg. Telecommun. Technol.* **2008**, *19*, 333–354, doi:10.1002/ett.1287. [CrossRef]

18. Kao, D.T.H.; Sabharwal, A. Two-user interference channels with local views: On capacity regions of TDM-dominating policies. *IEEE Trans. Inf. Theory* **2013**, *59*, 7014–7040, doi:10.1109/TIT.2013.2274512. [CrossRef]

19. Yeh, S.Y.; Wang, I. Degrees of freedom of the bursty MIMO X channel without feedback. *arXiv* **2016**, arXiv:1611.07630.

20. Jafar, S.A.; Vishwanath, S. Generalized degrees of freedom of the symmetric Gaussian *K* user interference channel. *IEEE Trans. Inf. Theory* **2010**, *56*, 3297–3303, doi:10.1109/TIT.2010.2048494. [CrossRef]

21. Tse, D.; Viswanath, P. *Fundamentals of Wireless Communication*; Cambridge University Press: Cambridge, UK, 2005.

22. Han, T.S.; Kobayashi, K. A new achievable rate region for the interference channel. *IEEE Trans. Inf. Theory* **1981**, *27*, 49–60, doi:10.1109/TIT.1981.1056307. [CrossRef]

23. Wang, I.H.; Suh, C.; Diggavi, S.; Viswanath, P. Bursty Interference Channel with Feedback. 2013. Available online: https://sites.google.com/site/ihsiangw/isit13preprintburstyic (accessed on 8 November 2018).

24. Sato, H. Two-user communication channels. *IEEE Trans. Inf. Theory* **1977**, *23*, 295–304. [CrossRef]

25. Suh, C.; Tse, D.N.C. Feedback capacity of the Gaussian interference channel to within 2 bits. *IEEE Trans. Inf. Theory* **2011**, *57*, 2667–2685, doi:10.1109/TIT.2011.2119990. [CrossRef]

26. Etkin, R.H.; Tse, D.N.C.; Wang, H. Gaussian interference channel capacity to within one bit. *IEEE Trans. Inf. Theory* **2008**, *54*, 5534–5562. doi:10.1109/TIT.2008.2006447. [CrossRef]

27. Verdú, S.; Han, T.S. A general formula for channel capacity. *IEEE Trans. Inf. Theory* **1994**, *40*, 1147–1157. doi:10.1109/18.335960. [CrossRef]

28. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley-Interscience: Hoboken, NJ, USA, 2006.

29. Rudin, W. *Real and Complex Analysis*, 3rd ed.; McGraw-Hill, Inc.: New York, NY, USA, 1987.